**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Uso de informações de parentesco e modelos mistos para avaliação e
seleção de genótipos de cana-de-açúcar**

Edjane Gonçalves de Freitas

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Genética e
Melhoramento de Plantas

**Piracicaba
2013**

Edjane Gonçalves de Freitas
Engenheira Agrônoma

**Uso de informações de parentesco e modelos mistos para avaliação e
seleção de genótipos de cana-de-açúcar**

Orientador:
Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Genética e
Melhoramento de Plantas

**Piracicaba
2013**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Freitas, Edjane Gonçalves de

Uso de informações de parentesco e modelos mistos para avaliação e seleção de genótipos de cana-de-açúcar / Edjane Gonçalves de Freitas.- - Piracicaba, 2013.
101p: il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2013.

1. Melhoramento genético 2. Seleção 3. Genealogias 4. Matriz de variâncias e covariâncias I. Título

CDD 633.61
F866u

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

"Quando alguém evolui,
também evolui tudo
ao seu redor..."

"Quando tentamos de ser
melhores do que somos, tudo
ao nosso redor também
se torna melhor."

"Você é livre para escolher...
Para tomar decisões...
Mesmo que só você consiga entender.
Tome-as com coragem...
Desprendimento e, às vezes,
com uma certa dose de loucura."

"Só entenderemos a vida
e o universo,
quando buscamos mais explicações.
Que fique esclarecido..."

"Aprender algo significa entrar em contato
com um mundo desconhecido,
onde as coisas simples,
são as mais extraordinárias.

Atreva-se a mudar.
Procure desafios...
Não tenha medo.
Insista uma... Insista outra...
E mais outra vez."

"Lembre-se com fé,
É possível ganhar uma batalha que parece perdida.
Não se dê por vencido, Tenha sempre a certeza do que quer..
E comece de novo."

"O segredo está em não ter medo de errar
e saber que é necessário
ser humilde para aprender."

"Tenha paciência para saber o momento certo
e comemore com os outros
e se isso ainda não for suficiente...
Analise as causas
e tente novamente com mais força..."

"O mundo está nas mãos
daqueles que têm a coragem
de sonhar e correr o risco
de viver seus sonhos..."

AGRADECIMENTOS

Nessa etapa final do doutorado penso quantas pessoas foram essenciais para que eu pudesse chegar até aqui. Esse sonho só foi possível graças a contribuição de cada um de vocês. Nunca esqueci do dia que deixei a minha família em Alagoas... Foi muito triste... Não fazia ideia do que iria enfrentar. O sonho de ter uma melhor condição vida foi decisivo para traçar esse caminho. Desde menina sempre acreditei que a educação era a única oportunidade de vencer. Alimentei esse pensamento dia após dia, e isso me deu forças e coragem para seguir em frente. Tive perdas, conquistas, mas sempre acreditei que valeria a pena. Durante esses anos muitos desafios enfrentei, e quando pensei que iria fraquejar sempre encontrei um amigo disposto para a ajudar. Por isso quero aproveitar esse momento para agradecer meu eterno orientador e grande amigo Prof. Geraldo Veríssimo, meu exemplo de pessoa e profissional. A o Prof. Hermann Hoffmann pelo acolhimento e carinho, e pela a oportunidade de fazer parte da equipe do PMGCA-UFSCar, sem a qual seria impossível a conquista do meu mestrado e continuar com o doutorado.

Ao Prof. Decio Barbin, foi uma alegria ter sido sua orientada no mestrado. Muito obrigada pelo apoio e confiança depositada em mim.

Ao Prof. Antonio Augusto, pela oportunidade e orientação no doutorado.

Ao prof. Roland Vencovsky que me recebeu em sua sala com muita atenção e educação. Nossa conversa foi esclarecedora. Seus questionamentos sobre o trabalho guiaram meu raciocínio.

A Dr^a Luciana Rossini, Dr. Mauro Xavier, Dr. Marco Landell pela concessão dos dados experimentais e pela disponibilidade para esclarecer dúvidas.

Um agradecimento especial reservo a minha querida amiga Maria Marta, uma pessoa admirável e digna. Desde o início nunca mediu esforços para ajudar. Sempre disposta a compartilhar conhecimento. Muito Obrigada Maria!!!

As minhas grandes amigas Rafinha e Paulinha. Durante esses anos, nos muitos momentos de angústias, sempre me ouviram e consolaram.

Ao admirável Rodrigo Gazaffi que no momento mais crítico, me ouviu e foi decisivo para o desfecho final desse trabalho. Sempre admirei você, mas hoje digo: “Esse é o cara”.

Ao Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo, pela oportunidade.

Ao CNPQ, pela concessão da bolsa e demais recursos ao longo desse trabalho.

Aos amigos do Laboratório de Genética Estatística: Maria Marta (a amiga), Rodrigo (o cara), Marcelo (cabeção), Gabriel (cabeção2), Graciela (um doce), Renato (o ortodoxo), João

Ricardo (o correto), Carina (a caridosa), Luciano (chefe), Guilherme (coração de ouro), Adriana (a poderosa), Rodrigo Júnior (tem futuro) e Rafael (o gente fina), pela amizade e conhecimento compartilhado .

Aos meus anjos da guarda Guilherme e Rafael, por tornarem o laboratório um ambiente mais receptivo e acolhedor. Pelo carinho, conforto, amizade e companheirismo. Por fazer os meus dias em Piracicaba mais felizes. Adoro vocês!!!

Ao João Ricardo, o “Jones”, que no último minuto do segundo tempo, ajudou na formatação da tese. Valeu Jones!!

Aos professores do Departamento de Genética e de Estatística da Escola Superior de Agricultura “Luiz de Queiroz”, pelos ensinamentos compartilhados.

Aos amigos e colegas do curso, pela convivência e aprendizado.

Aos funcionários do Departamento de Genética da ESALQ/USP: Seu Zé, Seu Antônio, Valdir, Berdan, Léia, Macedônio e Fernandinho, pela convivência durante todos esses anos.

A minha grande doce companheira e amiga Michi, sempre presente me confortando nas horas mais difíceis.

Um agradecimento especial reservo ao meu companheiro e grande amor Erick Espinoza. Você mostrou que a vida é muito mais do que 100 páginas escritas. Me consolou e tornou a minha vida mais leve e feliz. Hoje sei que embora as coisas não tenham saído como planejei, quando olho para você, vejo que apesar de tudo passei valeu a pena.

A todos amigos e familiares que direta ou indiretamente contribuíram para que esse trabalho fosse realizado. Muito Obrigado!

SUMÁRIO

RESUMO	9
ABSTRACT	11
1 INTRODUÇÃO	13
2 REVISÃO BIBLIOGRÁFICA	17
2.1 Cana-de-açúcar: importância econômica e melhoramento genético	17
2.2 Interação entre Genótipos e Ambientes e Grupos de Experimentos	18
2.2.1 Modelo Misto	22
2.2.2 Estimção em Modelos Mistos	25
2.2.2.1 Estimção dos Efeitos Fixos de β e Predição dos Efeitos Aleatórios de g	25
2.2.2.2 Estimção dos Parâmetros de Variâncias e Covariâncias em G e R	28
2.2.3 Modelo Misto para Análise Conjunta de Experimentos	30
2.2.4 Estruturas Alternativas de Matriz de Variâncias e Covariâncias para G^L e R_0	31
2.2.5 Incorporação de Parentesco Genético	35
2.2.6 Modelo para Seleção de Genótipos Superiores	38
2.2.7 Critérios para Seleção de Modelos de Variâncias e Covariâncias	39
3 MATERIAL E MÉTODOS	43
3.1 Material	43
3.1.1 Rede de experimentos	43
3.1.2 Material Vegetal	47
3.1.3 Caráter Quantitativo	48
3.2 Métodos	49
3.2.1 Matriz de Parentesco	49
3.2.2 Modelo Misto	50
4 RESULTADOS	57
4.1 Análise Usando Modelo Fixo	57
4.2 Coeficiente de Parentesco	59
4.3 Seleção do Modelo	64
5 DISCUSSÃO	79
6 CONCLUSÃO	83
REFERÊNCIAS	85
APÊNDICE	97

RESUMO

Uso de informações de parentesco e modelos mistos para avaliação e seleção de genótipos de cana-de-açúcar

Nos programas de melhoramento de cana-de-açúcar todos os anos são instalados experimentos com o objetivo de avaliar genótipos que podem eventualmente ser recomendados para o plantio, ou mesmo como genitores. Este objetivo é atingido com o emprego de experimentos em diferentes locais, durante diferentes colheitas. Além disso, frequentemente há grande desbalanceamento, pois nem todos os genótipos são avaliados em todos os experimentos. O emprego de abordagens tradicionais como análise de variância conjunta (ANAVA) é inviável devido à condição de desbalanceamento e ao fato de as pressuposições não modelarem adequadamente o relacionamento entre as observações. O emprego de modelos misto utilizando a metodologia REML/BLUP é uma alternativa para análise desses experimentos em cana-de-açúcar, permitindo também incorporar a informação de parentesco entre os indivíduos. Nesse contexto, foram analisados 44 experimentos (locais) de cana-de-açúcar do programa de melhoramento da cana-de-açúcar do Instituto Agronômico de Campinas (IAC), com 74 genótipos (clones e variedades) e com até 5 colheitas. O delineamento foi o de blocos ao acaso com 2 a 6 repetições. O caráter analisado foi TPH (Tonelada de pol por hectare). Foram testados 40 modelos, os 20 primeiros foram avaliadas diferentes estrutura de VCOV para locais e colheitas, e os 20 seguintes, além das matrizes de VCOV, foi incorporada a matriz de parentesco genético, A . De acordo com AIC, verificou-se que o Modelo 11, o qual assume as matrizes $FA1$, $AR1$ e ID , para locais, colheitas e genótipos, respectivamente, foi o melhor, e portanto, o mais eficiente para seleção de genótipos superiores. Quando comparado ao modelo tradicional (médias dos experimentos), houve mudanças no ranqueamento dos genótipos. Há correlação entre o modelo tradicional e o Modelo 11 ($\rho = 0,63$, $p\text{-valor} < 0,001$). A opção de utilizar modelo misto sem ajustar as matrizes de VCOV (Modelo 1) é relativamente melhor do que usar o Modelo Tradicional. Isto foi evidenciado pela correlação mais alta entre os modelos 1 e 11 ($\rho = 0,87$ com $p\text{-valor} < 0,001$). Acredita-se que o emprego do Modelo 11 junto com experiência do melhorista poderá aumentar a eficiência de seleção em programas de melhoramento de cana-de-açúcar.

Palavras-chave: Melhoramento genético; Seleção; Genealogias; Matriz de variâncias e covariâncias

ABSTRACT

Usage of kinship and mixed models for evaluation and selection of sugarcane genotypes

In breeding programs of sugarcane every year experiments are installed to evaluate the performance of genotypes, in order to select superior varieties and genitors. The use of ordinary approaches such as joint analysis of variance (ANOVA) is unfeasible due to unbalancing and assumptions that do not reflect the standard of relationship of the observations. The use of mixed models using the method REML/BLUP is an alternative. It also allows the incorporation of information from kinship between individuals. In this context, we analyzed 44 trials (locations) of sugarcane breeding program of sugarcane (Agronomic Institute Campinas, IAC), with 74 genotypes (varieties and clones), up to 5 harvests. The experimental design was randomized blocks with 2-6 replicates. The character was examined TPH (Tons of pol per hectare). We tested 40 models, the first 20 were evaluated different VCOV structure to locations and harvests, and 20 following addition of matrix VCOV was incorporated genetic relationship matrix, **A**. Under AIC, it was found that the model 11, which assumes matrices FA1, AR1 and ID for locations, harvests and genotypes, respectively, was the best. There is a moderate correlation between traditional model and model 11 ($\rho = 0.63$, p -value < 0.001), when ranking the genotypes. The option of using mixed model without adjusting matrices VCOV (model 1) is better than using the traditional model. This was suggested by the higher correlation between models 1 and 11 ($\rho = 0.87$ with p -value < 0.001). We believe that the usage of model 11 together with breeders experience can increase the efficiency of selection in sugarcane breeding programs.

Keywords: Breeding; Selection; Pedigree; Variance-Covariance Matrix

1 INTRODUÇÃO

A produção de biocombustíveis, fonte de energia renovável derivada de biomassa de cana-de-açúcar, diante da atual crise energética mundial, tornou-se uma alternativa promissora ao uso de combustíveis fósseis. No mundo, cresce o interesse pela exploração de biomassa de cana-de-açúcar, e o Brasil é destaque nesse segmento. Rico em matéria-prima, o país ocupa a posição de líder mundial na produção de cana-de-açúcar e seus derivados, com previsão do total de cana moída para safra 2012/2013 de 596,63 milhões de toneladas, sendo 50,42% destinado a produção de açúcar e 49,58% para a produção de etanol (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2012).

As vantagens ambientais e econômicas do biocombustível de cana criou um cenário favorável à pesquisa, e impulsionou o desenvolvimento e investimentos na área agrícola, permitindo através do melhoramento genético a criação de novas variedades de cana-de-açúcar altamente produtivas e com reduzido custo de produção. O sucesso da atividade sucroalcooleira no Brasil, em grande parte, é devido às pesquisas de melhoramento genético. O uso de novas metodologias, associadas aos avanços tecnológicos e científicos, tem fornecido novas ferramentas para aumentar a eficiência da seleção nos programas de melhoramento.

Um procedimento estatístico bastante utilizado para análise de experimentos de seleção e competição de variedades é a análise de variância (ANAVA). No modelo tradicional, todos os efeitos são considerados fixos, exceto o erro experimental, que é tratado como aleatório. Essa abordagem deu grandes contribuições ao melhoramento genético. Entretanto, é limitada por várias razões. Por exemplo, devido à pressuposição de independência dos erros, que no contexto dos experimentos de melhoramento genético pode ser inadequada. Outra limitação dessa técnica ocorre quando se tem desbalanceamento de dados, seja por perda de parcelas (comum em experimentos de campo), ou quando o delineamento não é ortogonal. Há também a situação em que cada parcela fornece vários dados em diferentes locais e anos (colheitas); por isso os dados são correlacionados. A estratégia usual consiste em realizar a transformação dos dados, mas isso pode não funcionar em muitos casos.

A eficiência de seleção num programa de melhoramento pode ser aumentada com o emprego de modelos estatísticos que representem mais fielmente a natureza dos experimentos. Isso pode ser conseguido com modelagem de estruturas de variância-covariâncias (VCOV) entre indivíduos, fornecida pela matriz de parentesco genético, A , e dentro de indivíduos nos diversos locais, dada pela interação genótipo $\times$ ambiente. As informações de covariância entre e dentro

de indivíduos, quando modelada adequadamente, aumenta o poder preditivo do modelo, e leva à obtenção de estimativas acuradas do valor genético (VA) e genotípico (VG). Consequentemente, aumenta a eficiência de seleção.

A metodologia de modelos mistos é uma alternativa aos modelos de ANAVA, tendo sido idealizada por Henderson, em 1984, em pesquisa de melhoramento genético animal. Permite-se modelar efeitos fixos e aleatórios além do erro experimental; é flexível, no sentido de permitir estruturar a informação de covariância, e tem sido aplicada com sucesso no melhoramento animal e de plantas. Peculiaridades, tais como a modelagem de informação de VCOV genética entre os efeitos do modelo e a incorporação de parentesco genético, despertam o interesse de pesquisadores, e a aplicação dessa abordagem tem aumentado, especialmente em pesquisas de Seleção Assistida por Marcadores (MAS) e Seleção Genômica Ampla (GWS)(CROSSA et al., 2006, 2007; BOER et al., 2007; BAUER; LÉON, 2008; LORENZANA; BERNARDO, 2009; SCHULZ-STREECK; PIEPHO, 2010; ZHIWU-ZHANG et al., 2010; BURGUEÑO et al., 2012). Fica claro que sua aplicação pode trazer muitas vantagens à experimentação com cana-de-açúcar.

O desenvolvimento de novas variedades de cana-de-açúcar, como também a liberação e a recomendação para plantio comercial requerem que etapas de seleção sejam praticadas entre um grande número de genótipos candidatos. É comum a obtenção de banco de dados extremamente desbalanceados em função de sucessivo descarte de genótipos ocorrido durante o processo de seleção e/ou implantação de ensaios com apenas alguns genótipos em comum. Nesses casos, a análise de variância conjunta tradicional torna-se difícil, ou mesmo impraticável. A análise conjunta dos experimentos com predição do valor genético aditivo e valor genotípico dos genótipos em experimentação é importante para estabelecer, durante o processo de seleção, comparações entre indivíduos em função da superioridade genética, e com efeito, auxiliar na decisão de seleção. Assim, o modelo misto juntamente com a metodologia REML/BLUP (do inglês, *Restricted Maximum Likelihood/Best Linear Unbiased Prediction* - REML/BLUP) pode fornecer predições mais acuradas do valor genético, medida comumente usada na identificação de genitores para futuros cruzamentos; e do valor genotípico, usado como referência no ranqueamento e seleção de genótipos superiores destinados à liberação comercial. Para o melhoramento de cana-de-açúcar, tal abordagem é uma boa alternativa que pode aumentar a eficiência de seleção. Quando um modelo é mais realista, este possui maior capacidade preditiva, conferida em função da inclusão de informação genealógica e modelagem da matriz de VCOV dos efeitos de interação genótipo $\times$ local $\times$ colheita ($G \times L \times C$). Adicionalmente, esse modelo explora a correlação entre as observações e permite a obtenção de predições dos genótipos com e sem informações

fenotípica, o que é de grande interesse para os programa de melhoramento.

Os experimentos com cana-de-açúcar normalmente são conduzidos em vários locais e anos (colheitas), podendo serem chamados de METs (do inglês, *multi-environment trial* - MET), fazendo parte de estratégia que visa aumentar a eficiência de seleção. Entretanto, modelos de predição mais realistas são necessários, incluindo simultaneamente a matriz de parentesco e uma adequada estrutura de variância-covariância para efeitos de interação. Algumas pesquisas de melhoramento genético de cana-de-açúcar tem reportado o uso do modelos mistos (RESENDE; BARBOSA, 2006; OLIVEIRA et al., 2008; ATKIN, DIETERS; STRINGER, 2009; PASTINA et al., 2012). Todavia, esses estudos consideram apenas separadamente a incorporação da informação de parentesco, ou a covariância dos efeitos de interação $G \times L \times C$.

Diante do exposto, considerando as características dos ensaios com cana-de-açúcar, o uso dos modelos mistos com incorporação da informação de parentesco e modelagem da matriz de variância-covariância para os efeitos de interação pode ser considerada uma estratégia promissora. Seu uso pode resultar em modelos com maior poder preditivo, fornecendo predições mais acuradas dos valores genético e genotípico. Tais modelos poderão auxiliar os melhoristas na tomada de decisão durante processo de seleção de genitores para futuros cruzamentos, bem como na recomendação de novas variedades para plantio comercial. Assim, essa abordagem foi utilizada na presente tese, usando dados reais fornecidos pelo programa de melhoramento do Centro da Cana do Instituto Agronômico de Campinas (IAC).

Nesse contexto, o objetivo desse trabalho foi determinar modelos para predição do valor genético de genótipos de cana-de-açúcar utilizando a abordagem de modelos mistos, com incorporação da matriz de parentesco, e modelando-se adequadamente a matriz de variância-covariância genética da interação $G \times L \times C$.

2 REVISÃO BIBLIOGRÁFICA

2.1 Cana-de-açúcar: importância econômica e melhoramento genético

A cana-de-açúcar (*Saccharum* spp.) é uma das principais culturas do Brasil e do mundo, com grande importância econômica, agroindustrial e social, gerando milhares de empregos diretos e indiretos. Já há alguns anos, o Brasil mantém a posição de maior produtor mundial de cana-de-açúcar e seus derivados. O setor sucroalcooleiro segue aquecido, acumulando recordes de produção. A previsão de safra 2012/2013 é de 596,63 milhões de toneladas de cana moída, sendo 50,42% destinado a produção de açúcar e 49,58% para a produção de etanol. A estimativa de área colhida e destinada à atividade foi de 8.527,770 mil hectares. A produtividade média brasileira foi estimada em 69.963 kg/ha. O estado de São Paulo é a principal região produtora, com 54% da produção nacional estimada (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2012).

A evidente expansão do setor sucroalcooleiro do Brasil tem enfrentado vários desafios, e o melhoramento genético tem sido essencial para assegurar o sucesso do setor. Todos os anos, os programas desenvolvem e liberam novas variedades com potencial produtivo superior às variedades comerciais. A introdução no mercado de materiais competitivos encara o desafio de atender a crescente demanda nacional e internacional no que concerne à produção de açúcar e álcool. Cresce a perspectiva de produção do etanol de segunda geração (oriundo da palha e/ou do bagaço do cana) em escala industrial e, paralelamente, aumenta-se o investimento em pesquisas (MOORE, 2005; CARDONA et al., 2010; RABELO, 2011; BUCKERIDGE et al., 2012). Nesse cenário positivo, o melhoramento genético busca aprimorar novas metodologias de análise (STRINGER; CULLIS, 2002; BARBOSA et al., 2004; ATKIN; DIETERS; STRINGER, 2009; BLANCO et al., 2010; PASTINA et al., 2012) a fim de otimizar o processo de obtenção e seleção de genótipos superiores, na expectativa de atender a exigência do mercado de açúcar e biocombustível, através da oferta de materiais genéticos com elevados índices de rendimento industrial.

A criação e a avaliação de variedades melhoradas é uma das mais antigas atividades da pesquisa agrícola, com grande impacto sobre a produção mundial de alimentos, especialmente nos últimos 100 anos (SMITH; CULLIS; THOMPSON, 2005). Especificamente no caso da cana-de-açúcar, na década de 1970 intensificaram-se as pesquisas de melhoramento genético no Brasil, e surgiram as primeiras variedades liberadas que atenderam o mercado de açúcar e álcool

(BARBOSA et al., 2012). Historicamente, o emprego do conhecimento de genética quantitativa foi marcante no melhoramento de cana, destacando-se os estudos de fatores genéticos relacionados especialmente ao potencial produtivo e ao entendimento e exploração da variabilidade genética (BROWN et al., 1968; HOGARTH et al., 1981; KANG, et al., 1983; MILLIGAN et al., 1990; JACKSON, 2005). Borlaug (2001) apud Smith, Cullis e Thompson (2005), argumenta que é importante reconhecer que os métodos convencionais de melhoramento continuam a fazer contribuições significantes para produção de alimentos, e, nesse sentido, é essencial que métodos estatísticos usados para analisar dados experimentais sejam precisos, eficientes e informativos quanto possível.

Nos programas de melhoramento de cana-de-açúcar, todos os anos são instalados vários experimentos com o objetivo de selecionar genótipos superiores, provenientes de hibridações entre genitores previamente selecionados (MARIOTTI, 1973; MATSUOKA et al., 1999). São vários genótipos candidatos, tornando a seleção mais difícil, e exigindo do melhorista o uso de ferramentas de análise apropriadas para aumentar a eficiência do processo de seleção e garantir o sucesso do programa de melhoramento. O emprego de métodos de genética quantitativa, junto com métodos estatísticos adequados, pode resultar em melhores estimativas de componentes variância genéticos e residuais, e consequentemente em predições mais acuradas do valor genético. Em essência, os programas de melhoramento de cana-de-açúcar baseia-se na seleção e clonagem genótipos superiores de populações segregantes, obtidas por meio de cruzamentos sexuais entre indivíduos diferente. Para maximizar a eficiência desse processo, são realizadas diferentes etapas, envolvendo a escolha adequada dos genitores e a quantificação dos efeitos ambientais na expressão de cada carácter sob seleção (MATSUOKA et al., 1999). E por ser uma espécie que permite a propagação clonal, toda a variabilidade genética pode ser explorada (SOUZA Jr., 1989; SOUZA Jr., 1995).

2.2 Interação entre Genótipos e Ambientes e Grupos de Experimentos

Tipicamente, os dados gerados no melhoramento de plantas são oriundos de uma série de ensaios estabelecidos em vários ambientes, exsquema conhecido como MET (SMITH; CULLIS; THOMPSON, 2001). Essa estratégia permite avaliar o desempenho de genótipos em vários ambientes com diferentes condições edafoclimáticas e em diferentes anos. Tradicionalmente, a análise de MET baseia-se em modelos simples, assume homogeneidade de variâncias e ausência de correlação entre as observações (SMITH et al., 2001; BALZARINI, 2001; SMITH; CULLIS; THOMPSON, 2005; SMITH et al., 2007). Dessa forma, o efeito de interação genótipo $\times$ am-

biente ($G \times E$, em que E é a combinação de local e anos), nesse modelo, assume uma matriz de VCOV do tipo independente (RESENDE; THOMPSON, 2004; SMITH; CULLIS; THOMPSON, 2005), ou seja, a variância genética dos ambientes é homogênea e não há correlação genética entre pares de ambientes. Claramente, tais suposições podem não ser corretas em muitos cenários.

O desenvolvimento de métodos para a análise de MET iniciou com método da ANAVA conjunta (KEMPTON, 1984; BALZARINI, 2001; SMITH et al., 2005). Quando aplicada para análise de rede de experimentos de cana-de-açúcar, particiona a variação total em fontes de variação devido a genótipo (G), a local (L), a colheita (C) e a interação $G \times L \times C$. Comparando métodos para análise de MET, Kempton (1984) apontou a desvantagem da ANAVA por não fornecer a introspecção sobre a natureza da estrutura de VCOV do efeito de interação $G \times L \times C$. Isso pode dificultar a seleção de genótipos e as decisões de recomendação. Pode ser insuficiente obter apenas a estimativa da média do desempenho de genótipos nos ambientes, e outras análises, tais como de estabilidade dos genótipos, são necessárias para identificação de variedades de alto desempenho produtivo e estáveis (apropriadas para uso amplo) e de bom desempenho apenas em certas condições (adequadas para utilização em ambiente específico). Técnicas de regressão linear (FINLAY, WILKINSON, 1963; EBERHART, RUSSELL; 1966) foram propostas e usadas no estudo da interação $G \times E$, porém são ineficientes na falta de linearidade (CROSSA, 1990; DUARTE; VENCOSKY, 1999). Alternativamente, sugeriu-se o emprego do modelo AMMI (do inglês, *additive main effects and multiplicative interaction analysis*) descrita por Gauch (1988; 1992) e atribuída a Fisher e Mackenzie (1923) e Gollob (1968). Essa técnica combina componentes aditivos para os efeitos principais de genótipos e ambientes, e componentes multiplicativos para o efeito de interação $G \times E$. Entretanto, possui a limitação de considerar os efeitos de genótipos e de interação como fixos (DUARTE; VENCOSKY, 1999), o que impossibilita a incorporação de informação de parentesco entre os indivíduos.

Nesse contexto, o emprego de modelos lineares mistos é uma alternativa aos modelos de efeitos fixos para análise de MET (PIEPHO, 1997; RESENDE; THOMPSON, 2004; SMITH; CULLIS; THOMPSON, 2005). Entre os argumentos a favor, ressaltam-se as vantagens de análise de dados desbalanceados (quando nem todos os genótipos estão em todos ambientes), situação real encontrada em um programa de melhoramento de cana-de-açúcar; a capacidade de usar modelos mais realistas do erro experimental (por exemplo, em blocos incompletos, modelar a correlação espacial); a possibilidade de admitir efeitos, por exemplo, de genótipos e/ou ambientes, como aleatórios ao invés de fixos, acomodando estrutura de correlação atribuída aos efeitos aleatórios. Essa abordagem tornou-se popular na análise de MET, com aplicações

variadas desde modelos simples de componentes de variância, que fornecem informações semelhantes a ANAVA, até modelos mistos multiplicativos, que visam explorar e acomodar melhor os efeitos de interação (BALZARINI, 2001; SMITH; CULLIS; THOMPSON, 2001; RESENDE; THOMPSON, 2004; SMITH; CULLIS; THOMPSON, 2005).

A presença de interação $G \times E$ é uma das dificuldades encontradas pelos melhoristas durante o processo de seleção. A expressão de um genótipo é alterada pelas condições ambientais (MEYER, 2009). Isso pode resultar em diferente ranqueamento de genótipos nos diversos ambientes. Desse modo, quando a avaliação ocorrer apenas em um local, a interação $G \times E$ pode inflacionar a estimativa de variância genética e gerar estimativas viesadas do ganho genético de seleção (ganhos reais inferiores ao previsto). Por outro lado, a implantação e análise de MET permite isolar o efeito de interação (DUARTE; VENCOVSKY, 1999). Quando a interação é complexa, os melhores genótipos num determinado local podem não ser em outros locais, ocorrendo mudança de ranqueamento e aumentando a dificuldade de seleção e recomendação do genótipo para todos ambientes que foram testados. Estatisticamente, isso decorre da impossibilidade de interpretar, de forma aditiva, os efeitos principais de genótipos e ambientes (DUARTE; VENCOVSKY, 1999). O efeito da interação é devido a heterogeneidade de variância genética entre os ambientes e a falta de correção perfeita de genótipos entre os pares de ambiente (FALCONER, 1952; BERNARDO, 2010). Assim, os efeitos principais de genótipo e ambiente não devem ser considerados separadamente. Cooper e DeLacy (1994) afirmaram que estudos do impacto da interação $G \times E$ em resposta à seleção deve distinguir entre esses dois componentes, para investigar a presença de heterogeneidade de variância dos genótipos e a correlação genética entre os ambientes. Entretanto, no contexto de modelos mistos, independente da natureza simples ou complexa da interação, esta pode ser modelada por uma matriz de variâncias e covariâncias, representada por G (PIEPHO, 1997, 1998, 2009; SMITH; CULLIS; THOMPSON, 2001; CROSSA et al., 2004, 2006; OAKLEY et al., 2006; BURGUEÑO et al., 2011, 2012).

Por outro lado, a estrutura do erro experimental em análise de MET no melhoramento genético é muito mais complexa que usualmente considerada em modelos lineares simples, reforçando ainda mais que o uso de modelos lineares tradicionais são inadequados, já que as pressuposições são irrealistas ao assumirem que os erros são não correlacionados e as variâncias são homogêneas. Normalmente, uma vez que isso não ocorra, os dados são transformados, ao invés de definir um novo modelo mais realista e mais coerente com o contexto de melhoramento genético (BALZARINI, 2001).

Os experimentos realizados no melhoramento da cana-de-açúcar são um caso típico de MET, pois todos os anos são instalados vários experimentos em diversas unidades produtoras distribuí-

das em diferentes regiões do país (SMITH et al., 2007). Além disso, genótipos são avaliados em diferentes locais e colheitas (planta, soca e ressoca), e, ao longo do processo, ocorre vários descartes, perdas de parcelas, que condicionam aos experimentos diferentes intensidades de desbalanceamento. Como, em geral, o interesse é avaliar a performance produtiva dos genótipos, estudar o comportamento, a adaptação e a estabilidade dos genótipos frente a diferentes condições de climas e solo, é fundamental explorar o efeito das interações genótipo $\times$ local ($G \times L$) e/ou genótipo $\times$ local $\times$ colheita ($G \times L \times C$). Portanto, o emprego da ANAVA não é adequado (KEMPTON, 1984; SMITH; CULLIS; THOMPSON, 2005). Primeiro, por se tratar de experimentos extremamente desbalanceados, algo difícil ou impossível de se fazer com as técnicas de ANAVA com modelos fixos. Segundo, mesmo que os dados sejam balanceados, esta abordagem não admite a existência de correlação genética entre as diferentes observações nas combinações de locais e anos (colheitas), e, também, não é possível modelar a estrutura de VCOV da interação $G \times L \times C$. No cenário de experimentação de cana-de-açúcar, é intuitivo admitir a existência de estrutura de VCOV para o efeito de interação diferente do modelo de independência imposta pelo modelo de ANAVA (SMITH; CULLIS; THOMPSON, 2005; SMITH et al., 2007). Um cenário comum consiste em se forçar artificialmente o balanceamento dos experimentos. A análise é realizada apenas com os genótipos (clones e variedades) comuns a todos experimentos, desconsiderando os demais. Dessa forma, é realizada a análise apenas dos genótipos comuns, descartando-se quantidade substancial dos dados, e em outra situação, não necessariamente melhor, é calculada a média aritmética de todos os genótipos, que em seguida, são ranqueadas. Obviamente estas não são as melhores soluções. Com o cenário real de desbalanceamento dos experimentos de cana-de-açúcar, é evidente que o emprego da ANAVA é inadequado e deve ser evitado. Nesse contexto, o uso de modelos lineares mistos é uma boa alternativa.

Falconer e Mackay (1996) enfatizaram que um carácter medido em dois diferentes ambientes não deve ser considerado como um único carácter. Essa percepção reflete a importância de investigar a interação $G \times E$. A ideia de caracteres correlacionados sugere uma forma alternativa para modelar os efeitos de interação (LYNCH; WALSH, 1998). Estendido a mais de dois locais, esse pensamento pode ser reforçado no contexto de modelo misto multivariado. Assim, modelar a estrutura de VCOV genética é possível e essencial para aumentar o poder preditivo de modelo de seleção. Em pesquisa, Burgueño et al. (2011) mostraram que o poder de predição de um modelo aumentou até 6% quando se modelou adequadamente a estrutura de VCOV para o efeito de interação $G \times E$. Resultados similares foram encontrados por So e Edwards (2011).

2.2.1 Modelo Misto

Os modelos lineares mistos foram propostos inicialmente por Henderson (1984) em pesquisa com melhoramento animal. Atualmente, fazem parte da rotina de análise desses programas. No melhoramento genético de plantas tem aplicação relativamente menor, mas tem tornado-se popular para análise de grupos de ensaios de variedades, METs (SMITH; CULLIS; THOMPSON, 2005; SMITH et al., 2007). Gradativamente, têm maior uso nos estudos de mapeamento de QTL (*Quantitative trait loci*) (PASTINA, 2010; MARGARIDO, 2011), mapeamento associativo (BRESEGHELLO; SORRELLS, 2006), seleção assistida (LANDE; THOMPSON, 1990; BOHN et al., 2001; DEKKERS; HOSPITAL, 2002; FLINT-GARCIA et al., 2003; BONNETT et al., 2005; COLLARD; MACKILL, 2008; HOSPITAL, 2009) e seleção genômica (MEUWISSEN, 2001; BRUMMER, 2004; YU et al., 2006; BERNARDO, 2007; RESENDE, 2008; CROSSA et al., 2010; ZHE-ZHANG et al., 2010; ZHIWU-ZHANG et al., 2010; HUANG et al., 2010; JANNINK; LORENZ; IWATA, 2010; SEGURA et al., 2012), com aplicações também no melhoramento florestal e em outras culturas de importância econômica.

O objetivo de um modelo estatístico é explicar as respostas de uma variável dependente em função de diferentes fatores que são atribuídos a uma série de variáveis independentes (SEARLE et al., 1992). De forma geral, no modelo linear misto, os efeitos dos fatores podem ser classificados como fixos ou aleatórios. Quando um determinado fator é atribuído a um conjunto finito de tratamentos ou níveis específicos de fatores experimentais, tal fator é de efeito fixo, e as conclusões, neste caso, são restritas aos níveis dos fatores (SEARLE et al., 1992; McCULLOCH; SEARLE, 2001). Por outro lado, quando um conjunto infinito de tratamento é atribuído ao fator, de tal forma que o conjunto seja uma amostra aleatória de uma população, o fator é dito de efeito aleatório, e é possível inferir sobre a população da qual os níveis do fator foram amostrados.

De acordo com a natureza fixa ou aleatória dos efeitos de um modelo, este é classificado como fixo, aleatório ou misto. Considera-se fixo, quando todos os seus efeitos são fixos, exceto o erro experimental, que é sempre aleatório. Considera-se o modelo como aleatório quando todos os efeitos são aleatórios, exceto a média geral (intercepto), que é sempre fixa. Há também a situação na qual se tem uma combinação de efeitos fixos e aleatórios, com exceção da média geral e do erro experimental, sendo definido como modelo misto (SEARLE et al., 1992; McCULLOCH; SEARLE, 2001). Segundo os autores, os dois tipos de efeitos (fixo e aleatório) são distintos e, por isso, na análise de dados, são tratados de forma distinta. Assim, a decisão de considerar um determinado efeito como fixo ou aleatório é crucial, e não é imediatamente óbvia chegar a tal decisão. Algumas considerações devem ser relevantes, tais como: a natureza dos dados, a forma

de coleta e o ambiente no qual os mesmos foram obtidos. Classicamente, o argumento utilizado para auxiliar na decisão sobre a natureza dos efeitos de um modelo fundamenta-se no fato que os níveis de um fator são amostrados de uma grande população (efeitos aleatórios) ou se foram definidos de acordo com o interesse do pesquisador (efeitos fixos) (SEARLE, 1971). Contudo, Smith, Cullis e Thompson (2005) afirmaram que essa escolha depende do objetivo da análise e de considerações a respeito das propriedades dos dois tipos de procedimentos de estimação, isto é, a predição linear não viesada, para os efeitos aleatórios, e a estimação linear não viesada, para os efeitos fixos. Pastina (2010) exemplifica com a situação típica de melhoramento, a seleção de variedades, em que o objetivo é ranquear os efeitos das estimativas dos genótipos o mais próximo possível do *rank* verdadeiro. Nesse caso, as estimativas dos efeitos de variedades devem prever os efeitos verdadeiros da melhor forma possível, o que implica no uso de *predição*, de tal forma que o efeito de variedade deve ser considerado aleatório. Nesse sentido, Burgueño et al. (2012) mostraram que a probabilidade de ocorrer uma alteração no ranqueamento de genótipos está associada ao coeficiente de regressão de Pearson (r), e quando a correlação entre os efeitos estimados e os efeitos verdadeiros é igual a um, a probabilidade de ocorrer uma alteração no ranqueamento é zero. Entretanto, quando o objetivo é determinar diferenças entre pares específicos de genótipos, Pastina (2010) acrescentou que o método de predição é inadequado, pois a predição de uma diferença específica é viesada. Nesse caso, o efeito de variedade deve ser considerado como fixo. Vale destacar que o uso de efeitos aleatórios de variedades ou genótipos apresenta a vantagem de permitir análises de conjunto de dados históricos combinados ao longo de vários anos (SMITH; CULLIS; THOMPSON, 2005).

É interessante mencionar que, frequentemente, a aplicação de modelos mistos no melhoramento de plantas tem dado ênfase a estimação de componentes de variância e identificação apropriada do erro experimental para testar as hipóteses dos efeitos fixos. E, raramente, tem sido usado uma proposta geral com a modelagem de estrutura de VCOV genética e predições dos efeitos aleatórios (BALZARINI, 2001).

Uma forma geral de modelo linear misto para análise de MET pode ser representada matricialmente como (HENDERSON, 1984; THOMPSON et al., 2003):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}$$

em que $\mathbf{Y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_J)'$ é o vetor de respostas fenotípicas, tomadas em I indivíduos ou genótipos $i = (1, 2, \dots, I)$ avaliados em J locais ($j = 1, 2, \dots, J$); $\mathbf{X}$ e $\mathbf{Z}$ são matrizes de delineamentos de posto completo, associadas ao vetor $\boldsymbol{\beta}$ de parâmetros fixos, $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_J)'$ e o

vetor $\mathbf{g} = (\mathbf{g}'_1, \mathbf{g}'_2, \dots, \mathbf{g}'_J)'$ de efeito genético aleatório, respectivamente; e $\boldsymbol{\varepsilon} = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_J)'$ é o vetor de erro aleatório. Os vetores aleatórios, $\mathbf{g}$ e $\boldsymbol{\varepsilon}$, assumem distribuição normal com média zero, ou seja, $E(\mathbf{g}) = 0$ e $E(\boldsymbol{\varepsilon}) = 0$, e estruturas de variâncias e covariâncias (VCOV) expressadas da seguinte forma:

$$\begin{bmatrix} \mathbf{g} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right),$$

com as matrizes de VCOV para o vetor de resíduo $\boldsymbol{\varepsilon}$ dada por $\mathbf{R}$ e de efeitos genéticos aleatórios expressa por $\mathbf{G}$. Assim desde que $E(\mathbf{g}) = 0$ e $E(\boldsymbol{\varepsilon}) = 0$ por definição $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$.

Desta forma, tem-se que:

$$\mathbf{V} = Var(\mathbf{Y}) = Var(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R},$$

em que o primeiro termo explica a contribuição dos efeitos genéticos aleatórios, enquanto o segundo apresenta a variância devido ao efeito residual. Caso o procedimento usual de análise seja aplicado, assume-se que a variância residual é constante (homogeneidade de variância) e não correlacionada; isto é, o modelo tradicional de ANAVA. Então, $\mathbf{R}$ neste caso é uma matriz diagonal, com $\mathbf{R} = \sigma_e^2 \mathbf{I}$.

Assumindo ainda que $\mathbf{V}$ é não singular,

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta}$$

implica que,

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})$$

No modelo linear misto para análise de MET, a matriz de VCOV genética pode ser decomposta, de forma que $\mathbf{G} = \mathbf{G}^L \otimes \mathbf{A}$, em que $\mathbf{G}^L$ é a matriz $J \times J$ de informação de covariância genética de genótipos avaliadas nos diferentes locais, atribuída ao efeito genético principal e de interação $G \times L$; $\otimes$ denota o produto de Kronecker (ou produto direto) entre as duas matrizes; e $\mathbf{A} = \{a(i, i')\}$ é uma matriz $I \times I$ de parentesco genético aditivo, conhecida como *numerator relationship matrix*, e seus elementos são duas vezes o coeficiente de parentesco ou coancestria, $2\Phi_{i, i'}$, entre os pares de indivíduos. Ainda, a $Cov(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}') = \mathbf{R} = \mathbf{I}_I \otimes \mathbf{R}_0$ e a $Cov(\mathbf{g}, \boldsymbol{\varepsilon}) = \mathbf{0}$. A matriz de covariância global $\mathbf{G}$, dada por $Cov(\mathbf{g}, \mathbf{g}') = \mathbf{G}^L \otimes \mathbf{A}$ é comumente representada por (BURGUEÑO et al., 2012):

$$\mathbf{G}^L \otimes \mathbf{A} = \begin{bmatrix} \sigma_{a_1}^2 & \rho_{12}\sigma_{a_1}\sigma_{a_2} & \cdots & \rho_{1J}\sigma_{a_1}\sigma_{a_J} \\ \rho_{21}\sigma_{a_1}\sigma_{a_2} & \sigma_{a_2}^2 & \cdots & . \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{J1}\sigma_{a_J}\sigma_{a_1} & . & \cdots & \sigma_{a_J}^2 \end{bmatrix} \otimes \mathbf{A},$$

em que o j -ésimo elemento da diagonal da matriz $\mathbf{G}^L$ é a variância genética aditiva $\sigma_{a_j}^2$ no j -ésimo local, sendo o elemento $j'j$ a covariância genética aditiva ($\rho_{j'j}\sigma_{a_{j'}}\sigma_{a_j}$) entre os locais j' e j ; assim, $\rho_{j'j}$ é a correlação dos efeitos genéticos aditivo entre locais j' e j .

A matriz de VCOV residual $\mathbf{R}_0 = \{Cov(\varepsilon_{ij}, \varepsilon_{ij'})\}$ tem dimensão $J \times J$ e modela a correlação espacial entre parcelas. É possível acomodar a heterogeneidade de variância residual substituindo $\mathbf{I}_I$ do produto direto, $\mathbf{I}_I \otimes \mathbf{R}_0$, por uma matriz diagonal, $\mathbf{D} = Diag\{\sigma_j^2\}$.

Com essas pressuposições, e seguindo as propriedades de distribuição normal multivariada, a densidade marginal dos dados é normal multivariada e pode ser expressa por:

$$[\mathbf{Y}|\boldsymbol{\beta}, \mathbf{R}_0, \mathbf{G}^L] \sim \mathbf{N}_{MV}[\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}(\mathbf{G}^L \otimes \mathbf{A})\mathbf{Z}' + \mathbf{I}_I \otimes \mathbf{R}_0]$$

As estimativas dos parâmetros, $\hat{\boldsymbol{\beta}}$, e a predição dos efeitos genéticos aleatórios, $\hat{\mathbf{g}}$, são as soluções do sistema de equações do modelo misto de acordo com Henderson (1984).

2.2.2 Estimação em Modelos Mistos

Para estimar os parâmetros e predizer os efeitos aleatórios, a estrutura especial da distribuição normal para $\mathbf{g}$ e $\boldsymbol{\varepsilon}$ permite dividir o desenvolvimento teórico em duas etapas. Na primeira, consideram-se $\mathbf{G}$ e $\mathbf{R}$ conhecidas, obtendo-se a solução para $\mathbf{g}$ e $\boldsymbol{\beta}$. Depois, estimam-se os parâmetros não conhecidos em $\mathbf{G}$ e $\mathbf{R}$ (método iterativo).

2.2.2.1 Estimação dos Efeitos Fixos de $\boldsymbol{\beta}$ e Predição dos Efeitos Aleatórios de $\mathbf{g}$

A obtenção do sistema de equações normais para os modelos mistos pode ser feita pela minimização da soma de quadrados dos resíduos ou pela maximização da função de densidade de probabilidade conjunta de $\mathbf{Y}$ e $\mathbf{g}$ (LITTEL et al., 2006).

As deduções apresentadas a seguir baseiam-se em Littel et al. (2006); Henderson (1984); Thompson et al. (2003).

A função densidade de probabilidade de $\mathbf{Y}$ é dada por:

$$f(\mathbf{y}) = \frac{1}{2\pi^{\frac{n}{2}}(\mathbf{ZGZ}' + \mathbf{R})^{\frac{1}{2}}} e^{-\frac{1}{2}[(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{ZGZ}' + \mathbf{R})^{-1}(\mathbf{y}-\mathbf{X}\beta)]}.$$

A função densidade de probabilidade conjunta de $\mathbf{Y}$ e $\mathbf{g}$ pode ser escrita como produto entre a função densidade condicional de $\mathbf{Y}$ dado $\mathbf{g}$, e a função densidade de probabilidade de $\mathbf{g}$, conforme segue:

$$f(\mathbf{Y}, \mathbf{g}) = f(\mathbf{Y}|\mathbf{g}) \cdot f(\mathbf{g})$$

$$f(\mathbf{Y}, \mathbf{g}) = \frac{1}{2\pi^{\frac{n}{2}}|\mathbf{R}|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\mathbf{y}-\mathbf{X}\beta-\mathbf{Zg})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\beta-\mathbf{Zg})]} \cdot \frac{1}{2\pi^{\frac{q}{2}}|\mathbf{G}|^{\frac{1}{2}}} e^{-\frac{1}{2}[(\mathbf{g}-0)'\mathbf{G}^{-1}(\mathbf{g}-0)]}$$

sendo $|\mathbf{G}|$ e $|\mathbf{R}|$ os determinantes das matrizes de VCOV.

Para proceder a maximização de $f(\mathbf{Y}, \mathbf{g})$, pode-se usar a transformação por logaritmo. Assim, o logaritmo da função de verossimilhança $L(\beta, \mathbf{g}|\mathbf{Y})$ é

$$l(\beta, \mathbf{g}|\mathbf{Y}) = \frac{1}{2}2n \log(2\pi) - \frac{1}{2}(\log |\mathbf{R}| + \log |\mathbf{G}|) - \frac{1}{2}(\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Y} - 2\mathbf{Y}'\mathbf{R}^{-1}\mathbf{X}\beta - 2\mathbf{Y}'\mathbf{R}^{-1}\mathbf{Zg} + 2\beta'\mathbf{X}'\mathbf{R}^{-1}\mathbf{Zg} + \beta'\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\beta + \mathbf{g}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Zg} + \mathbf{g}'\mathbf{G}^{-1}\mathbf{g}).$$

Derivando-se $l(\beta, \mathbf{g}|\mathbf{Y})$ em relação a β e a $\mathbf{g}$, tornando-se tais derivadas identicamente nulas, obtém-se:

$$\begin{bmatrix} \frac{\partial l(\beta, \mathbf{g}|\mathbf{Y})}{\partial \beta} \\ \frac{\partial l(\beta, \mathbf{g}|\mathbf{Y})}{\partial \mathbf{g}} \end{bmatrix} = \begin{bmatrix} -\mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{g}} \\ -\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{g}} + \mathbf{G}^{-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{g}} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\hat{\beta} + \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\hat{\mathbf{g}} + \mathbf{G}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix},$$

e assim,

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix} \quad (1)$$

As equações de modelos mistos (EMM) de Henderson, permitem obter conjuntamente as soluções para os efeitos fixos $\hat{\beta}$ e as predições dos efeitos genéticos aleatórios $\hat{\mathbf{g}}$. As EMM são equações normais estendidas ou também equações dos quadrados mínimos generalizados. A estimação é mais complexa no modelo misto do que no modelo linear geral, pois além de $\hat{\beta}$, agora o modelo tem parâmetros desconhecidos em $\mathbf{g}$, $\mathbf{G}$ e $\mathbf{R}$. Neste caso, o método de mínimos quadrados não é o melhor, sendo o método de mínimos quadrados generalizado (do inglês, *generalized least squares* - GLS), mais apropriado, minimizando $(\mathbf{Y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$.

As soluções são dadas por:

$$\begin{bmatrix} \hat{\beta} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix},$$

sendo $\mathbf{X}^{-}$ uma inversa generalizada de $\mathbf{X}$.

Para os efeitos fixos, têm-se

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y},$$

que é o estimador de mínimos quadrados generalizados. Verifica-se que, $\hat{\beta}$, das EMM, é uma solução GLS para o modelo $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, que ignora o efeitos aleatórios, g .

O preditor de g é o Melhor Preditor Linear Não Viesado (BLUP). O termo preditor refere-se a fatores aleatórios, e o BLUP pode ser, resumidamente, definido como o resultado da regressão dos efeitos de um fator aleatório g em função das observações $\mathbf{Y}$ corrigidas para os efeitos dos fatores fixos, $\mathbf{X}\beta$. Portanto, os BLUPs do efeito genético g é obtido conforme a seguinte expressão:

$$\hat{g} = \mathbf{GZ}'\mathbf{V}^{-1}[\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}] = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\beta),$$

em que o termo $\mathbf{GZ}'\mathbf{V}^{-1}$ é o conjunto de coeficientes de regressão de g em função $\mathbf{Y}$, uma vez que $\mathbf{GZ}'$ é a matriz de covariâncias entre g e $\mathbf{Y}$; e $\mathbf{V}^{-1}$ é a inversa da matriz de VCOV de $\mathbf{Y}$, já definida anteriormente. O termo $(\mathbf{Y} - \mathbf{X}\beta)$, contém os valores das observações, $\mathbf{Y}$, corrigidas para os efeitos fixos $\mathbf{X}\beta$.

Como nas EMM, g pode ser dado por

$$\hat{g} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\beta),$$

assim, se a igualdade

$$\mathbf{GZ}'\mathbf{V}^{-1} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1},$$

for verdadeira, $\hat{g}$, obtido pela EMM, é o BLUP de g . A prova dessa igualdade foi apresentada por Henderson et al. (1959).

Em resumo, o estimador de β e o preditor de g , são respectivamente, o estimador de mínimos quadrados generalizados (GLSE) ou o melhor estimador linear não enviesado (BLUE) e o melhor preditor linear não enviesado (BLUP)(SEARLE et al., 1992). Teoricamente, o BLUP com a

abordagem de modelos mistos, tem propriedades importantes para melhoramento genético vegetal, entre elas: a) em um único procedimento, permite a estimação e predição não enviesadas; b) pode considerar os efeitos de seleção e endogamia ao longo das gerações, desde que o grau de relacionamento genético entre os indivíduos avaliados seja conhecido; c) maximiza a correlação entre os valores genéticos verdadeiros e os preditos (SEARLE et al., 1992), de grande interesse para o melhorista; d) pode prever o valor genético de genótipos com e sem informação fenotípica; e) as variâncias e os erros das predições BLUP são menores em relação a outros métodos; f) possui o menor erro quadrático médio dentre todos os preditores lineares não enviesados (WHITE; HODGE, 1989).

Henderson et al. (1959) argumentam que em caso de experimentos balanceados, conhecendo ou não o parentesco entre os genótipos, as soluções do modelo misto (BLUP) e as soluções dos quadrados mínimos (BLUE) podem levar ao mesmo ordenamento dos genótipos. E, para que BLUE e BLUP detenham a propriedade de mínimo erro quadrático médio, os componentes de variância genéticos G e não genéticos R devem ser conhecidos.

2.2.2.2 Estimação dos Parâmetros de Variâncias e Covariâncias em G e R

Na prática, a matriz de variâncias e covariâncias V é desconhecida, assim como G e R também são, sendo necessário obter suas estimativas por algum método (ROBINSON, 1991). As estimativas de G e R podem ser obtidas através de métodos de Máxima Verossimilhança (ML, do inglês *Maximum Likelihood*) e Máxima Verossimilhança Restrita (REML, do inglês *Restricted Maximum Likelihood*), aplicados a $(Y - X\beta)'V^{-1}(Y - X\beta)$. Diferentemente dos estimadores da ANAVA, o ML e REML não possuem nenhuma exigência sobre o delineamento e/ou desbalanceamento dos dados. O método que tem sido rotineiramente utilizado para estimar componentes de variância em modelos mistos é o REML, desenvolvido por Patterson e Thompson (1971). Ele é uma variante que elimina o viés do método ML das estimativa de componentes de variância (PATTERSON; THOMPSON, 1971). Assim, de acordo com Resende (2004), o BLUP é o procedimento ótimo de predição de valores genéticos e o REML é o procedimento ótimo de estimação de componentes de variância.

A estimação dos parâmetros desconhecidos ocorre através da maximização de uma função objeto em relação a G e R . Assim para o método da Máxima Verossimilhança (ML), o logaritmo da função de verossimilhança é expresso como (PATTERSON; THOMPSON, 1971; ROBINSON, 1991; LITTEL et al., 2006):

$$l_{ML}(G, R) = -\frac{1}{2} \log |V| - \frac{n}{2} \log [Y - X(X'V^{-1}X)X'V^{-1}Y]'V^{-1}[Y - X(X'V^{-1}X)X'V^{-1}Y]$$

$$-\frac{n}{2} \left[1 + \log\left(\frac{2\pi}{n}\right) \right].$$

Enquanto o método da Máxima Verossimilhança Restrita (REML) possui logaritmo da função de verossimilhança expressa por:

$$l_{REML}(\mathbf{G}, \mathbf{R}) = -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$$

$$-\frac{n-p}{2} \log [\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}]' \mathbf{V}^{-1} [\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}]$$

$$-\frac{n-p}{2} \log \left[1 + \log \left(\frac{2\pi}{n-p} \right) \right]$$

em que p é o posto da matriz $\mathbf{X}$.

Nos dois caso, a solução das expressões requer o emprego de algoritmos iterativos tais como o *NR - Newton-Rapson* (YPMA, 1995), *Fisher Scoring* (PATTERSON; THOMPSON, 1971), o *EM - Expectation Maximization* (DEMPSTER; LAIRD; RUBIN, 1977) e o *AI - Average Information* (implementado no *software* ASREML e embutido no GenStat (2011)). O algoritmo *AI* junto com o método de matriz esparsa é indicado para ajuste de modelos mistos, permitindo a análise de grande e complexo conjunto de dados com rapidez e eficiência (GILMOUR et al., 2009).

Estimadas as matrizes $\hat{\mathbf{G}}$ e $\hat{\mathbf{R}}$, estas substituem as matrizes $\mathbf{G}$ e $\mathbf{R}$ na expressão dada em (1). Então, a estimação de β e predição de $\mathbf{g}$, são obtidas por

$$\begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Y} \end{bmatrix} = \hat{\mathbf{C}} \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Y} \end{bmatrix},$$

em que $\hat{\mathbf{C}}$ é a matriz de VCOV estimada dos erros de estimação e predição ($\hat{\beta} - \beta, \hat{\mathbf{g}} - \mathbf{g}$) das equações de modelos mistos (LITTELL et al., 2006). McLean e Sanders (1988) mostraram que $\hat{\mathbf{C}}$ pode ser escrita como

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{\mathbf{C}}_{11} & \hat{\mathbf{C}}'_{21} \\ \hat{\mathbf{C}}_{21} & \hat{\mathbf{C}}_{22} \end{bmatrix}$$

Com

$$\hat{\mathbf{C}}_{11} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$$

$$\hat{\mathbf{C}}_{21} = -\hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}\mathbf{X}\hat{\mathbf{C}}_{11}$$

$$\hat{\mathbf{C}}_{22} = (\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1})^{-1} - \hat{\mathbf{C}}_{21}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Z}\hat{\mathbf{G}}$$

As submatrizes, $\hat{C}_{11}$, $\hat{C}_{21}$ $\hat{C}_{22}$ têm, respectivamente, dimensões, $p \times p$, $p \times q$ e $q \times q$. Usando essa notação, Henderson (1975) mostrou que a matriz de VCOV de β , com $Var(\hat{\beta} - \beta) = Var(\hat{\beta})$, uma vez que $Var(\beta) = 0$, pois β é fixo. Assim, resulta que $Var(\hat{\beta} - \beta) = Var(\hat{\beta}) = \hat{C}_{11}$, a qual é a fórmula usual da variância das estimativas de quadrados mínimos generalizados. Da mesma forma, a matriz de VCOV do erro de predição de $\hat{g} - g$, é obtida por $Var(\hat{g} - g) = \hat{C}_{22}$ e, finalmente, a VCOV do erro de predição de $\hat{\beta}$ e $\hat{g} - g$, é dada por $Var(\hat{\beta}, \hat{g} - g) = \hat{C}_{21}$.

Contudo, quando G e R são substituídas por suas estimativas, $\hat{G}$ e $\hat{R}$, os termos BLUE e BLUP não mais se aplicam, e o termo *Empírico* ou *Estimado* é acrescentado para indicar essa aproximação. Agora, têm-se novos acrônimos, o *E-BLUE* (*empirical best linear unbiased estimator*) e o *E-BLUP* (*empirical best linear unbiased predictor*), para o BLUE e BLUP, respectivamente (LITTEL et al., 1996).

Os *E-BLUPs* do valor genotípico dos genótipos, ($E-BLUP(g)$), são menores que os valores genotípicos estimados por modelos fixos. Isso ocorre devido a ponderação da média por pesos que são funções da razão entre componentes de variância genéticos e ambientais. Por esse motivo, o BLUP dos efeitos aleatórios é comumente chamado de estimador *shrinkage* ou de “encolhimento” (BALZARINI, 2001). É natural que surja o questionamento a respeito de ordenamento de médias de genótipos quando considera um efeito aleatório ao invés de fixo. Segundo Duarte e Vencovsky (2001), dado que somente o modelo misto utiliza a informação relativa às variabilidades genotípicas das populações, é possível surgir classificação distinta entre os dois enfoques.

2.2.3 Modelo Misto para Análise Conjunta de Experimentos

O modelo para análise de grupos de experimentos pode ser representado conforme segue (RESENDE; THOMPSON, 2004):

$$y_{ij} = \mu + g_i + l_j + gl_{ij} + \varepsilon_{ij}$$

em que y_{ij} é a resposta fenotípica do genótipo i no ambiente j ; μ é a média geral; l_j é o efeito de local; g_i é o efeito genético do genótipo i ; gl_{ij} é efeito de interação $G \times L$; ε_{ij} é o erro aleatório. Aqui, μ e l_j podem ser consideradas fixos, e os demais aleatórios. No contexto de MET, um modelo com o efeito aleatório de genótipo em cada ambiente pode ser escrito como

$$y_{ij} = \mu + g_{ij} + l_j + \varepsilon_{ij}$$

em que g_{ij} é o efeito do genótipo i no local j . E, diferentes estruturas de VCOV, G^L para g_{ij} , podem ser consideradas (Tabela 1). Na matriz G^L , os elementos da diagonal são a variância genética em cada local (reflete a magnitude da variação entre os genótipos em cada local), e, fora da diagonal, as covariâncias genéticas entre os pares de locais (reflete a concordância de ordenamento de genótipos) (SMITH; CULLIS; THOMPSON, 2001).

No esquema de MET as matrizes covariâncias R_0 e G^L podem ser do tipo não estruturadas (do inglês *Unstructured* - UNST) e contém $J(J - 1)/2$ parâmetros. Esse número cresce proporcional ao número de ambiente. Quando J é grande, os efeitos genéticos e/ou residual são altamente correlacionados entre os ambientes, ao passo que, a estimativa de R_0 ou G^L torna-se quase singular, e o processo de convergência é lento. O uso de diferentes estruturas de VCOV é a solução a esse impasse, e podem ser ajustadas para R_0 e/ou G^L (BURGUEÑO et al., 2012).

2.2.4 Estruturas Alternativas de Matriz de Variâncias e Covariâncias para G^L e R_0

Uma breve descrição de cada estrutura (Tabela 1) é necessária no intuito de orientar na seleção da estrutura adequada, a qual deve corresponder aos padrões de resposta das observações. A matriz G^L com 1 na diagonal é o modelo de independência (ID) que considera independência e homogeneidade de variância, e é o modelo assumido pela ANAVA tradicional. Possui pressuposições irrealistas no contexto de melhoramento genético, uma vez que não existe correlação genética entre os pares de ambientes (sustenta a hipótese de presença de interação $G \times E$ do tipo complexa) e há homogeneidade de variância genética dos ambientes. A falta de correlação genética entre locais indica que um genótipo responde diferentemente às variações ambientais, e ocorre alteração de sua ordenação nos ambientes. O mesmo raciocínio se aplica para avaliar demais modelos.

Na sequência, o modelo diagonal (DIAG), admite independência e heterogeneidade de variância genética entre ambientes; implica em assumir uma variância genética separada para cada ambiente e ausência de covariância genética entre os ambientes. Neste caso, os ambientes são não correlacionados, e isso é similar a análise de cada ambiente separadamente. Patterson et al. (1977) consideram o modelo Simetria Composta (CS) que assume que todos ambientes tem a mesma variância genética e todos os pares de ambientes tem a mesma covariância. O modelo de Patterson et al. (1977) não tenta modelar o efeito de interação, gerando informação apenas de sua magnitude. E, também, ignora a possibilidade de heterogeneidade de variância dos ambientes. Cullis et al. (1998) ajustaram o modelo de Simetria Composta Heterogênea (CS_{Het}) que assumem heterogeneidade de variância dos ambientes e a existência de mesma correlação

Tabela 1 – Modelos alternativos de estrutura de variâncias e covariâncias (VCOV) para a matriz G que podem ser consideradas na análise de MET

Modelo	Descrição	Matriz G
Identidade (ID)		$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & 0 & \cdots & 0 \\ 0 & \sigma_g^2 + \sigma_{ge}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
Diagonal (DIAG)		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$
Simetria Composta (CS)		$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \sigma_g^2 & \cdots & \sigma_g^2 \\ \sigma_g^2 & \sigma_g^2 + \sigma_{ge}^2 & \cdots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \cdots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
Simetria Composta Heterogênea (CS _{Het})		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_g^2 & \cdots & \sigma_g^2 \\ \sigma_g^2 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 & \sigma_g^2 & \cdots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$
Autoregressiva de 1ª Ordem (AR1)		$\begin{bmatrix} \sigma_g^2 + \sigma_{ge}^2 & \sigma_g^2 \rho_g & \cdots & \sigma_g^2 \rho_g^{d(1,J)} \\ \sigma_g^2 \rho_g & \sigma_g^2 + \sigma_{ge}^2 & \cdots & \sigma_g^2 \rho_g^{d(2,J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(J,1)} & \sigma_g^2 \rho_g^{d(J,2)} & \cdots & \sigma_g^2 + \sigma_{ge}^2 \end{bmatrix}$
Autoregressiva de 1ª Ordem heterogênea (AR1 _{Het})		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_g^2 \rho_g & \cdots & \sigma_g^2 \rho_g^{d(1,J)} \\ \sigma_g^2 \rho_g & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & \sigma_g^2 \rho_g^{d(2,J)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_g^2 \rho_g^{d(J,1)} & \sigma_g^2 \rho_g^{d(J,2)} & \cdots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$
Fator Analítico de 1ª Ordem (AF1)		$\begin{bmatrix} \lambda_1^2 + \Psi_1 & \lambda_1 \lambda_2 & \cdots & \lambda_1 \lambda_J \\ \lambda_2 \lambda_1 & \lambda_2^2 + \Psi_1 & \cdots & \lambda_2 \lambda_J \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_J \lambda_1 & \lambda_J \lambda_2 & \cdots & \lambda_J^2 + \Psi_J \end{bmatrix}$
Não Estruturada (UNST)		$\begin{bmatrix} \sigma_{g_1}^2 + \sigma_{ge_1}^2 & \sigma_{g_{12}}^2 & \cdots & \sigma_{g_{1J}}^2 \\ \sigma_{g_{21}}^2 & \sigma_{g_2}^2 + \sigma_{ge_2}^2 & \cdots & \sigma_g^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g_{J1}}^2 & \sigma_{g_{J2}}^2 & \cdots & \sigma_{g_J}^2 + \sigma_{ge_J}^2 \end{bmatrix}$

σ_g^2 e σ_{ge}^2 : componentes de variância para o efeito principal de genótipos e de interação genótipos $\times$ ambientes, respectivamente; $\sigma_g^2 \rho_g^{d(j,j')}$: correlação genética entre ambientes, em que $d(j, j')$ corresponde à distância em tempos entre eles; $\sigma_{g_j}^2$ e $\sigma_{ge_j}^2$: variância genética específica de cada ambiente para o efeito principal de genótipos e de interação genótipos $\times$ ambientes, respectivamente; $\sigma_{jj'}$: covariância genética entre os ambiente j e j' ; Ψ_j : variância residual específica de cada ambiente; λ_j e λ_j' : elementos (*loadings*) do fator nos ambientes j e j' . Os acrônimos referentes aos vários modelos foram mantidos na forma derivada do inglês e de acordo com os padrões do *software* GenStat.

genética entre os pares de ambientes, refletindo a concordância do ranqueamento de genótipos entre os ambientes.

A matriz auto-regressiva de primeira ordem (AR1) e heterogênea (AR1_{Het}) apresentam a pressuposição de correlação genética entre os ambientes, com variância homogênea e heterogênea, respectivamente. Pastina (2010) ressalta a importância desses modelos, especialmente para culturas perenes e semiperenes, tais como cana-de-açúcar, cujos experimentos normalmente são avaliados diversas colheitas (planta, soca e ressoça), em anos distintos. É um caso de medidas repetidas no tempo. Para essas culturas, Pastina (2010) acrescenta que a correlação genética pode diminuir com a distância temporal entre as colheitas, o que confere uma interessante interpretação para este tipo de modelo. Uma justificativa é que genes expressos na primeira colheita (cana planta) podem não ser expressos em anos subsequentes (PASTINA, 2010). Outro modelo é o fator analítico de primeira ordem (FA1), é uma aproximação do modelo não estruturado (UNST), com o número de fator $k = 1$. Muitas pesquisas tem sugerido o modelo fator analítico na análise de MET (PIEPHO, 1998; SMITH; CULLIS; THOMPSON, 2001; KELLY et al., 2007; SO; EDWARDS, 2011; BURGUEÑO et al., 2012). O modelo não estruturado assume a estrutura geral, em que $\mathbf{G}$ é completamente não estruturada com $J(J + 1)/2$ parâmetros para diferentes variâncias genéticas de cada ambiente e diferentes covariância genética entre os pares de ambientes. Entretanto, como já mencionado, quando o número de ambiente avaliado é grande, a estimativa dessa matriz é ineficiente ou não estimável, até mesmo para um número moderado de ambiente. Portanto, uma estrutura mais parcimoniosa é desejável (SMITH; CULLIS; THOMPSON, 2001).

Modelo do tipo Fator Analítico (FA), apresentados por Smith, Cullis e Thompson (2001), que geram uma aproximação do modelo não estruturado, sendo geralmente preferidos por fornecer acurácia da predição do E-BLUPs do valor genotípico, sendo comumente usado para modelar a matriz de VCOV em modelos de genética quantitativa (PIEPHO, 1997, 1998; SMITH; CULLIS; THOMPSON, 2001; RESENDE; THOMPSON, 2004; CROSSA et al., 2004, 2006; BURGUEÑO et al., 2008, 2011; 2012). Devido à sua importância, merece maior atenção, e maior enfoque metodológico será apresentado.

A estrutura FA pode ser vista como uma extensão da análise de componentes principais. No contexto de MET, o modelo de fator analítico de ordem k , FA k , pode ser usado para modelar a matriz de variâncias e covariâncias, $\mathbf{G}$, dos efeitos genéticos nos ambientes, g_{ij} . É postulado em termos de efeitos genotípicos em diferentes ambientes (RESENDE; THOMPSON, 2004;

PASTINA, 2010):

$$g_{ij} = \sum_{r=1}^k \lambda_{jr} f_{ir} + \delta_{ij},$$

em que g_{ij} é o efeito do genótipo i no ambiente j ; $\sum_{r=1}^k \lambda_{jr} f_{ir}$ é o somatório dos termos multiplicativo que explicam a interação $G \times L$; em que λ_{jr} é peso ou carga para o fator r (variável latente) no ambiente j ; f_{ir} é o *score* para o genótipo i no fator r , ou seja, o fator comum; e δ_{ij} é o erro devido a falta de ajuste do modelo, ou seja, o fator específico associado ao genótipo i no ambiente j .

Portanto, a matriz $\mathbf{G}$ é modelada por dois tipos de fatores, comuns e específicos. Dessa forma, se a matriz de VCOV for definida por $\mathbf{G} = \mathbf{G}^L \otimes \mathbf{A}$, decorre que $(\Lambda\Lambda' + \Psi) \otimes \mathbf{A}$, para $\text{FA}k$. Na Tabela 1, em $\mathbf{G}$, sob modelo FA , assume-se que $\sigma_{g,j}^2 + \sigma_{ge,j}^2 = \sum_{r=1}^k \lambda_{jr}^2 + \Psi_j$, representando a variância genética nos ambientes, em que Ψ_j é a variância genética residual ou o fator genético específico de δ_{ij} ; e $\sigma_{jj'}^2 = \sum_{r=1}^k \lambda_{jr} \lambda_{j'r}$ é a covariância entre os ambiente j e j' .

Quando apenas um fator $k = 1$ é considerado, o modelo é denotado como FA1 ; para $k = 2$, FA2 , indicando dois componentes multiplicativos. O modelo FA pode ser interpretado como um modelo de regressão linear do efeito de genótipo e de interação $G \times L$ sobre covariáveis ambientais (peso ou carga ambiental, λ). Cada genótipo tem um coeficiente angular específico (*score* genotípico) e um intercepto comum (se os efeitos principais de genótipos não forem distinguidos da interação $G \times L$). O coeficiente angular mede a sensibilidade dos genótipos aos fatores ambientais representados pelo “peso” de cada ambiente (SMITH; CULLIS; THOMPSON, 2001; BURGUEÑO, et al., 2012).

Burgueño et al. (2012), utilizando a abordagem de modelos mistos e assumindo a estrutura FA para a matriz $\mathbf{G}^L$, avaliaram linhagens de trigo em grupos de experimentos, e constataram que a eficiência de seleção em um programa de melhoramento pode aumentar com o emprego de modelos com maior poder preditivo. Isso é conseguido modelando-se adequadamente a matriz de variâncias e covariâncias genéticas para o efeito de interação $G \times L$, e incorporando-se as relações de parentescos genético entre os indivíduos.

Smith et al. (2007) propuseram um método para análise MET e seleção de indivíduos superiores de culturas perenes. Esse é um caso típico que ocorre em programa de melhoramento de cana-de-açúcar, em que a produção é obtida de sucessivas colheitas e de uma série de ensaios em diferentes locais, de forma que nem todos os genótipos são testados em todos experimentos. Os autores sugeriram um método alternativo para a estrutura de VCOV da matriz $\mathbf{G}$, que con-

siste numa extensão do abordagem de Smith; Cullis; Thompson (2001). Só que, ao invés de considerar uma única matriz VCOV ($\mathbf{G}$) para as combinações de locais e colheitas (em que combinação local-colheita considerada um ambiente), é assumido duas diferentes estruturas de VCOV genética, uma para local ($\mathbf{G}_{J \times J}^L$) e outra para colheita ($\mathbf{G}_{K \times K}^C$), além da matriz de genótipos ($\mathbf{I}_{I \times I}^G$), uma matriz identidade com I genótipos. Nesse contexto, a matriz de VCOV genética global $\mathbf{G}$ dos efeitos genéticos de colheitas e de locais é obtida por

$$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G, \quad (2)$$

em que $j = 1, 2, \dots, J$ é o número de locais de avaliação, $k = 1, 2, \dots, K$, é o número de colheitas. As formas separadas de VCOV para locais e colheitas são possíveis mesmo quando alguns locais tiverem menos que K colheitas. As estruturas apresentadas na Tabela 1 se aplicam a qualquer uma das duas matrizes. Obviamente, a escolha das estruturas a serem testadas deve ser coerente com o padrão de resposta das observações. Nesse sentido, para análise de MET de cana-de-açúcar, se justifica a tentativa de ajuste de matriz de VCOV do tipo DIAG, FA ou UNST para locais (SMITH et al., 2007), e as do tipo AR ou AR_{Het} para colheitas (PASTINA et al., 2012). Trata-se de um caso de medida repetidas, portanto, há uma variação temporal entre as observações tomadas nas diferentes colheitas, a qual precisa ser modelada.

Quando modela-se adequadamente estruturas de VCOV para matrizes $\mathbf{G}_{J \times J}^L$ e $\mathbf{G}_{K \times K}^C$, aumenta-se a acurácia da predição e o poder preditivo do modelo de seleção (SMITH et al., 2001; CROSSA et al., 2006; BURGUEÑO et al., 2008; KELLY et al., 2009). No entanto, melhores resultados poderão ser obtidos quando adicionalmente é modelada a correlação entre os genótipos aparentados (CROSSA et al., 2006; OAKEY et al., 2006; KELLY et al., 2009; BURGUEÑO et al., 2012). Apesar de pesquisas constatarem a contribuição do parentesco genético na melhoria de modelo de predição (WEI; BORRALHO, 2000; PURBA et al., 2001; FURLANI et al., 2005; PIEPHO et al., 2008; ATKIN; DIETERS; STRINGER, 2009; KELLY et al., 2009; BURGUEÑO et al., 2012), a literatura é escassa de aplicação de modelos mistos com incorporação de parentesco genético conjuntamente com a modelagem de matriz de VCOV genética, $\mathbf{G}$.

2.2.5 Incorporação de Parentesco Genético

O conhecimento da relação de parentesco entre os indivíduos é muito importante para os programas de melhoramento. O coeficiente de coancestria, Φ (também chamado coeficiente de *kinship*, de consanguinidade ou de parentesco) é uma medida clássica de parentesco genético. No melhoramento de plantas, Φ é útil para determinar o tamanho efetivo da população em culti-

vares alógamas (SOUZA; SORRELLS, 1989), seleção parental (COWEN; FREY, 1987), atribuir germoplasma a diferentes grupos heteróticos ou melhorados, e também para especificar a distância genética mínima para proteção varietal (HUNTER, 1989).

O parentesco genético é determinado com base em genealogias ou, preferencialmente, através de marcadores moleculares. Para tanto, é construída uma matriz de parentesco genético aditivo, **A**. Cada elemento dessa matriz é duas vezes o coeficiente de coancestria, $2\Phi_{i,i'}$, entre os pares de indivíduos. O coeficiente de coancestria $\Phi_{i,i'}$ dos indivíduos i e i' foi originalmente definido por Malécot (1969) como a probabilidade que dois gametas tomados aleatoriamente, um em cada indivíduo, carregarem alelos homólogos idênticos por descendência. Se S_{il} representa um alelo aleatoriamente amostrado de um indivíduo i e loco l , com uma similar definição para $S_{i'l}$, o coeficiente de coancestria entre dois indivíduos é definido por

$$\Phi_{i,i'l} = p(S_{il} \equiv S_{i'l}),$$

em que $\equiv$ denota idêntico por descendência.

O parentesco genético é comumente incorporado ao modelo através da matriz **A**. Contém toda a informação sobre o fluxo de genes na população, permitindo a explícita dissecação da variância genética (Van der WERF, 2011). Foi originalmente usada por Henderson (1984) no modelo animal, contendo a covariância entre os indivíduos relacionados, gerada por usar a informação dos ancestrais. Em modelos de predição do valor genético, o uso da matriz **A** indica que a informação de parentesco entre os indivíduos será usada.

Piepho et al. (2008) revisaram a aplicação de BLUP no melhoramento de plantas incluindo a informação de parentesco genético para explorar a correlação genética entre indivíduos aparentados. Ressaltaram que, quando o modelo considera essa informação, o BLUP do valor genético resulta em predições mais acuradas quando comparada com o BLUP sem a informação de parentesco. Faz sentido, portanto, considerar essa informação na avaliação de experimentos.

Portanto, a inclusão da matriz de parentesco no modelo de seleção junto com adequada estrutura de VCOV resulta em modelos mais realistas, com maior acurácia e capacidade preditiva. Burgueño et al. (2012) salientaram que o poder preditivo é maior e o ranqueamento é mais eficiente quando as duas informações (a matriz de VCOV genética mais a matriz **A**) são consideradas no modelo. Desse modo, a predição do valor genético retém informação entre indivíduos, fornecida pela matriz **A**, e dentro de indivíduos (entre os ambientes) através da matriz **G**, sendo um modelo mais realista, justificando a maior acurácia e maximização da eficiência de seleção. O uso da matriz de parentesco deve melhorar a precisão das estimativas dos valores genéticos e

quase sempre aumenta a correlação entre os valores genéticos dos indivíduos aparentados (HENDERSON, 1976; DEMPFLÉ, 1990). Diversos trabalhos comprovam que a inclusão da matriz de parentesco aditivo aumenta a acurácia das avaliações genéticas (KENNEDY; MOXLEY, 1975; POLLAK et al., 1977; CARLSON et al., 1984).

Estudos já evidenciaram que a incorporação da genealogia no modelo de seleção pode aumentar a acurácia da predição do valor genético, de forma a melhorar a eficiência de seleção e aumentar o ganho genético (BERNARDO et al., 1994; CROSSA et al., 2006; OAKEY et al., 2006; BURGUEÑO et al., 2007; ZHANG et al., 2010; KERR et al., 2012). A predição do valor genético baseada na genealogia fundamenta-se no modelo infinitesimal (FISHER, 1918) e tem desempenhado um papel central no melhoramento de caracteres complexos em plantas. O melhoramento animal tem utilizado essa abordagem para prever o valor genético com modelo misto (BLUP) (HENDERSON, 1984) ou o enfoque bayesiano (GIANOLA; FERNANDO, 1986).

Nesse sentido, modelos com incorporação de A e da matriz $VCOV$ poderá otimizar a decisão de seleção de parentais para futuros cruzamentos. O modelo com incorporação de parentesco fornece o BLUP do valor genético. Este é usado para medir o potencial de um indivíduo como genitor (WHITE; HODGE, 1989).

Em teoria, de acordo com Falconer e Mackay (1996) o valor genético (parte do valor genotípico), computa a informação de variância genética aditiva, fração do valor genotípico que é transmitida para descendência. Se um indivíduo é cruzado com outros tomados aleatoriamente de uma população, o seu valor genético é duas vezes a diferença entre a média da progênie e a média da população (a multiplicação por dois é devido ao genitor em questão fornece apenas metade dos genes para sua descendência, a outra metade vem aleatoriamente da população). Ainda segundo os autores, considerando um único loco, a diferença entre valor genotípico (VG) e valor genético (VA) de um particular genótipo é conhecido como desvio de dominância (VD), esses três efeitos podem ser relacionados por $VG = VA + VD$. O efeito de dominância surge da interação alélica no loco, e, na ausência de dominância, VA é igual a VG . Na situação em que mais de um loco é considerado, por exemplo, loco l e loco l' , surge uma variação adicional, a epistasia (VI), efeito de interação entre os locos originada pela ação combinada dos alelos de diferentes genes, conhecida como interação gênica. E, em função da forma como genes interagem, poderão modificar-se as proporções mendelianas. Nesse contexto, se VG_l é o valor genotípico do indivíduo considerando o loco l e $VG_{l'}$ o valor genotípico para o segundo loco, assim, VG , o valor genotípico admitindo os dois locos, é equacionado de forma que $VG = VG_l + VG_{l'} + VI_{ll'}$, em que $VI_{ll'}$ é o desvio da combinação aditiva do valor genotípico (VG). Assim, para todos

os locos, $VG = VA + VD + VI$, em que VA é agora a soma do valor genético de cada loco e VD a soma dos desvios de dominância. Esse conhecimento pode ser direcionado de acordo com os objetivos do programa de melhoramento. É importante para distinguir entre dois modelos: o modelo para seleção de genitores e o modelo para seleção de genótipos superiores para recomendação comercial.

Em termos práticos, considerando que espécie estudada é a cana-de-açúcar, a diferença entre valor genético e valor genotípico se resume no fato que o primeiro computa o efeito genético aditivo, portanto o modelo com essa informação é destinada a seleção de genitores. Já o segundo, computa os efeitos genéticos aditivo e não aditivo (dominância, nesse caso), sendo, portanto, usado para a seleção de genótipos superiores (OAKLEY et al., 2007).

2.2.6 Modelo para Seleção de Genótipos Superiores

A essência da seleção consiste em aumentar a frequência genes favoráveis e reduzir a frequência dos genes desfavoráveis. Nesse sentido, o ponto de partida do melhoramento genético é a definição de uma população base que disponibilize variabilidade genética; desta população, são selecionados genitores com características de interesse para o melhoramento cuja composição genética está associada, em geral, ao desempenho produtivo, a resistência as principais pragas e doenças e a tolerância ao estresse hídrico, especialmente nas regiões produtoras com maior deficit hídrico (MACKILL et al., 1999; SLAFER et al., 2005; TRETHOWAN et al., 2005).

Um procedimento comum em programas de melhoramento de cana-de-açúcar é o cruzamento entre genótipos superiores, muito dos quais são cultivares comerciais ou pré-comerciais que possuem conjunto gênico favorável previamente fixado. A escolha dos genitores e o planejamento dos cruzamentos são importantes etapas para o sucesso de um programa de melhoramento. O planejamento cuidadoso dos cruzamentos aumenta as chances de desenvolvimento de variedades superiores, pois maximiza a utilização de genes desejáveis (SOUZA Jr., 1995; ALLARD, 1999).

Desenvolver um bom modelo de predição do valor genotípico é fundamental. Ele ajudará a definir estratégia de seleção, contribuirá para construção de ranqueamento verdadeiro e, conseqüentemente, para a seleção dos genótipos superiores ou futuras variedades para liberação comercial. O sucesso das pesquisas de melhoramento genético de diferentes espécies de plantas depende basicamente da eficiência de criação e de identificação de genótipos superiores com qualidades agrônômicas. Nos dois casos, a seleção desempenha um papel fundamental e de grande importância, pois, quando bem sucedida, aumenta as chances de obtenção de uma nova

variedade.

A continuidade do melhoramento genético requer constantes estudos. No melhoramento da cana-de-açúcar, a liberação de uma nova variedade comercial leva de 11 a 15 anos, o que reforça a necessidade de otimizar todas as etapas que envolvem processo de seleção. Esse fato ressalta a importância de se ter preditores acurados de componentes de variância e consequentemente, do valor genotípico, uma vez que é essencial para aumentar a eficiência do processo de seleção. Nesse sentido, o uso de modelos que permitam estrutura a matriz de VCOV é uma boa alternativa, pois maximiza a correlação entre o valor genotípico verdadeiro e o predito, atendendo aos objetivos dos melhoristas (SEARLE et al., 1992).

Os procedimentos ótimos de predição do valor genéticos e seleção no melhoramento genético de plantas são índices multiefeitos que usam todos os efeitos aleatórios do modelo estatístico, ideal para casos balanceados, e o BLUP, indicado para os casos desbalanceados. Segundo Resende (2000), o BLUP ajusta os dados para os efeitos ambientais identificáveis, tais como efeito de bloco, de colheita e de local de plantio, e simultaneamente prediz os valores genéticos dos indivíduos. Ainda de acordo com a autor, as técnicas de estimação baseadas no método de quadrados mínimos, como a análise de variância, não são mais recomendadas para a aplicação no melhoramento de plantas perenes.

Os modelos mistos permitem conjuntamente incorporar a matriz de parentesco existente entre os genótipos de cana-de-açúcar e modelar mais adequadamente estruturas de VCOV para os efeitos de local e colheita. Outra vantagem dessa abordagem é sua aplicação a dados desbalanceados, seja pela perda de observações ou pelo delineamento, fato bastante comum em experimento de campo de cana-de-açúcar.

2.2.7 Critérios para Seleção de Modelos de Variâncias e Covariâncias

A seleção do modelo é um tópico complexo, na qual diversas estruturas de VCOV para matriz G podem ser razoáveis. Considerações sobre o problema em questão e objetivos devem estar presentes quando se seleciona um modelo. Depois da seleção de alguns modelos plausíveis, a escolha da estrutura de VCOV pode ser auxiliada por alguns métodos (DIGGLE, 1988; LINDSEY, 1999; BROW; PRESCOTT, 2006; VERBEKE; MOLENBERGHS, 2009). Várias estratégias tem sido adotadas para orientação, tais como métodos de diagnóstico, gráficos (CHRISTENSEN; PEARSON; JOHNSON, 1992), métodos fundamentados no princípios de verossimilhança (DIGGLE, 1988; OMAN, 1991; WOLFINGER, 1993). Dentre os métodos utilizados na seleção da estrutura de VCOV destacam-se

i) Teste Assintótico da Razão de Verossimilhança (LR)

$$LR = -2 \times \log \left(\frac{L_R}{L_C} \right), \quad (3)$$

em que L_R é a verossimilhança residual do modelo reduzido, e L_C a verossimilhança residual do modelo completo. Esse teste pode ser utilizado em modelos mistos aninhados. Consiste em comparar dois modelos estimados por máxima verossimilhança, em que um deles é uma versão restrita do outro, ou seja, um modelo completo tem r parâmetros adicionais. Assegurado que a parte fixa do modelo seja a mesma para ambos modelos, apenas o número de parâmetros da estrutura de variâncias e covariâncias é testado. O teste irá verificar se esses parâmetros melhoram significativamente o modelo. A hipótese testada (H_0) é a de que os dois modelos são equivalentes (os parâmetros extras não diferem de zero). A estatística resultante da diferença ($L_R - L_C$) tem distribuição assintótica Qui-Quadrado (χ_r^2) com r graus de liberdade (MOOD et al., 1974).

ii) Critério de Informação

Como alternativa ao teste de razão de verossimilhança, pode-se tomar uma perspectiva na teoria da informação e computar medidas baseadas na informação. Estas são calculadas como um termo de penalização aplicado a função de verossimilhança. Os mais comuns são o Critério de Informação de Akaike (AIC - AKAIKE, 1974) e o Critério de Informação Bayesiano (BIC - SCHWARZ, 1978) para comparação de modelos misto não aninhados, os quais são definidos como (BOZDOGAN, 1987; WOLFINGER, 1993):

$$AIC = -2 \log(L) + 2 \times n_{PAR} \quad (4)$$

$$BIC = -2 \log(L) + \log(N) \times n_{PAR} \quad (5)$$

em que $\log(L)$ é o logaritmo da função de máxima verossimilhança residual, n_{PAR} é o número total de parâmetros da matriz de VCOV e N é o número total de observações (HU; SPILKE, 2011). Menores valores desses critérios indica o melhor modelo, haja vista que o logaritmo da função de máxima verossimilhança tem coeficiente negativo nas duas expressões, de forma que a inclusão de parâmetros aumenta os valores de AIC e BIC. No entanto, AIC tenta escolher o modelo que minimiza a divergência de Kullback-Leibler (BURNHAM; ANDERSON, 2004). Essa divergência está relacionada à informação perdida por se usar um modelo aproximado e não o “real” (frequentemente é conceitualizada com uma “distância” entre modelo “real” completo e outro modelo) (PASTINA, 2010). Segundo Guerin e Stroup (2000), o AIC tende a selecionar

modelos mais complexos, que melhor se ajustam aos dados, mas, em compensação, com mais parâmetros. Além disso, possui maior controle da taxa de erro tipo I. O BIC, por sua vez, parte do pressuposto da existência de um “modelo verdadeiro”, maximizando a probabilidade para identificá-lo, penalizando mais fortemente o número de parâmetros do modelo. Em síntese, o LR, AIC e BIC são conceitualmente diferentes mas utilizam o mesmo critério estatístico, o máximo da função de verossimilhança como medida de ajuste. Entretanto, definem valores críticos diferentes (BOZDOGAN, 1987; WOLFINGER, 1993; LITTELL et al., 2002).

O cálculo desses critérios não envolve os parâmetros referentes à parte fixa do modelo. Desse modo, a significância desses parâmetros especificamente não são avaliadas por esses critérios. Assim, as inferências relativas aos parâmetros de efeitos fixos geralmente são testadas utilizando-se o teste de Wald (VERBEKE; MOLENBERGHS, 2009).

3 MATERIAL E MÉTODOS

3.1 Material

3.1.1 Rede de experimentos

O banco de dados utilizado para as análises foi gentilmente fornecido pelo o programa de melhoramento genético da cana-de-açúcar do Instituto Agronômico de Campinas - IAC. As informações são de um período de doze anos, referentes as colheitas realizadas de 1997 a 2001 e de 2003 a 2009, totalizando 74 genótipos (clones e variedades) avaliados em 44 experimentos. A rede experimental do IAC é composta por várias unidades de produção distribuídas em diferentes regiões canavieiras do estado de São Paulo, em Minas Gerais, Goiás e Mato Grosso, conforme mostrados na Figura 1 e na Tabela 2. Algumas unidades produtoras possuem mais de um experimento (Tabela 2). O número de cortes foi variável em cada experimento e será discutido oportunamente.

Figura 1 – Distribuição geográfica dos 44 experimentos analisados

A Figura 1 apresenta um panorama da distribuição geográfica dos experimentos considerados nesta tese. Há uma abrangência de distintos ambientes de produção. Trata-se de uma

Tabela 2 – Ano de plantio e colheita e localização dos 44 experimentos

Ano		Estado	Município	Usina	Experimento
Plantio	Colheita				
1996	1997-1999	MG	Delta	Us. Delta	5
1996	1997-1999	SP	Iguaraçu do Tietê	Us. Da Barra	2
1996	1997-1999	SP	Pradópolis	Us. São Martinho	14
1996	1997-1999	SP	Ribeirão Preto	Us. Galo Bravo	8
1996	1997-1999	SP	São João da Boa Vista	Us. Dedini	4
1996	1997,1999-2000	SP	São Joaquim da Barra	Us. Alta Mogiana	11
1996	1997-1998	SP	Sertãozinho	Us. Sta. Elisa	12
1996	1997-1998	SP	Sta. Rita do Passa Quatro	Us. Sta. Rita	13
1996	1997-1999	SP	Valparaíso	Us. Univalem	15
1996	1997-1999,2001	SP	Viradouro	Us. Virálcool	16
2002	2003-2006	SP	Catanduva	Us. Catanduva	3
2002	2003-2007	MG	Delta	Us. Delta	10
2002	2003-2006	GO	Goianésia	Us. Jalles Machado	17
2002	2003-2006	GO	Goiatuba	Us. Goiás	7
2002	2003-2005	SP	Promissão	Equipav	6
2002	2003-2006	SP	Sales Oliveira	Faz. Nova Aliança	9
2002	2004-2006	SP	São Joaquim da Barra	Us. Alta Mogiana	1
2003	2004-2006	SP	Araraquara	Us. Sta.Cruz O. P.	36
2003	2004-2006	SP	Catanduva	Us. Catanduva	20
2003	2004-2007	MG	Delta	Us. Delta	22
2003	2004-2007	GO	Goianésia	Us. Jalles Machado	30
2003	2004-2007	GO	Goiatuba	Us. Goiás	24
2003	2004-2007	SP	Guariba	Us. Bonfim	19
2003	2004-2007	SP	Motuca	Us. Sta. Luiza	37
2003	2004-2006	SP	Piracicaba	Us. Costa Pinto	21
2003	2004-2006	SP	Pradópolis	Us. São Martinho	43
2003	2004,2005	SP	Promissão	Equipav	23
2003	2004-2007	SP	Sales Oliveira	Faz. Nova Aliança	31
2003	2004-2007	SP	São Joaquim da Barra	Us. Alta Mogiana	18
2003	2004-2007	SP	Serrana	Us. Da Pedra	34
2003	2004-2006	SP	Severínia	Açúcar Guarani	26
2003	2004-2007	SP	Sta. Rita do Passa Quatro	Us. Sta. Rita	38
2004	2005-2007	SP	Araraquara	Us. Sta.Cruz O. P.	40
2004	2005-2008	SP	Araras	Us. São João	42
2004	2005-2008	GO	Goianésia	Us. Jalles Machado	29
2004	2005-2009	GO	Goiatuba	Us. Goiás	25
2004	2005-2007	SP	Motuca	Us. Sta. Luiza	39
2004	2005-2007	MT	Nova Olímpia	Us. Itamarati	28
2004	2005-2008	SP	Pradópolis	Us. São Martinho	44
2004	2005-2008	SP	Sales Oliveira	Faz. Nova Aliança	32
2004	2005-2008	SP	Serrana	Us. Da Pedra	35
2004	2005-2006,2008	SP	Severínia	Açúcar Guarani	27
2004	2005-2006	SP	Sta. Rita do Passa Quatro	Us. Sta. Rita	41
2004	2005-2006	SP	Tarumã	Us. Nova América	33

grande área sujeita a acentuada diferença nos parâmetros climáticos, com variação de temperatura, precipitação e intensidade luminosa. Essas variações afetam principalmente o período de crescimento vegetativo da cana-de-açúcar.

Grande parte dos experimentos está concentrada ao norte do estado de São Paulo, contem-

plando as classes de solos e climas mais representativos da indústria canavieira paulista. As coordenadas geográficas variam de 20°34'53" latitude sul e 47°51'17" longitude oeste (ao norte do estado de São Paulo), onde situa-se o município de São Joaquim da Barra, com clima tropical semi-úmido, temperatura variando de 15,9 a 29, 2°C e precipitação média anual de 1432,4 mm; estendendo-se até região oeste onde situa-se o município de Tarumã, com 22°44'48" latitude sul e 50°34'38" longitude oeste, com clima quente e úmido, temperatura média anual variando de 16 a 29, 8°C, precipitação média anual de 1430,4 mm. Há também, três experimentos que foram instalados no estado de Minas Gerais, no município Delta (19°58'36" latitude sul e 47°46'16" longitude oeste, clima subtropical úmido, temperatura média anual variando de 16,6 a 29, 1°C e precipitação média anual de 1.550 mm), seis em Goiás, sendo três em Goianésia (15°19'03" latitude sul e 49°07'03" longitude oeste, clima tropical úmido, temperatura média anual variando de 17 a 33°C e precipitação média anual de 1.700 mm) e três em Goiatuba (18°00'48" latitude sul e 49°21'30" longitude oeste, temperatura variando de 19 a 38°C, clima tropical quente e úmido e precipitação média anual de 1.500 mm), e um em Mato Grosso no município de Nova Olímpia (14°47'50" latitude sul e 57°17'17" longitude oeste, clima tropical quente e sub-úmido, temperatura variando de 0 a 40°C, precipitação média anual de 1.750 mm). Essas informações sustentam a hipótese que possivelmente existe heterogeneidade de variância entre os 44 experimentos avaliados, o que obviamente precisa ser considerado nas análises estatísticas. Pode-se esperar também que o comportamento relativo dos genótipos varie entre os ambientes, evidenciando a presença de interação $G \times E$.

Na presente tese, cada experimento analisado foi interpretado como um único ambiente ou local. Todos experimentos foram instalados usando o delineamento em blocos ao acaso, com o número de repetições variando de 3 a 6. Em geral, o tamanho da parcela foi de 5 sulcos de 8 metros e com espaçamento de 1,50 metros. Entre as unidades produtoras há uma variação de cerca de 5% no tamanho da parcela, em função especialmente do tipo de colheita (mecanizada ou manual), tratos culturais e manejo. Foram avaliadas de 1 a 5 colheitas (também chamadas de cortes), incluindo o primeiro ano (cana-planta), segundo ano (cana-soca) e até mais três anos seguidos de ressoca (Tabela 3). Essa condição extremamente desbalanceada reflete uma situação cotidiana enfrentada por muitos programas de melhoramento de cana do Brasil e, portanto, servirá como um ótimo e realista exemplo do cenário que pode ser considerado para as análises com o modelo proposto.

Tabela 3 – Número de colheitas de cada genótipo por experimento

(Continua)

[illegible]

Tabela 3 – Número de colheitas de cada genótipo por experimento

(Conclusão)

Genótipos	Experimentos																				Total
	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
IAC811032	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC822045	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC823092	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC831313	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC832285	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC832396	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC832405	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC834107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC853229	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC862210	5	0	0	0	0	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	47
IAC873184	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC873187	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
IAC873396	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	3	0	0	0	0	44
IAC911099	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IAC911121	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IAC911206	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP931020	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP932060	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP933046	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP933050	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP936035	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP942094	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP942101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP942111	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP942180	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP943581	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP943591	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP944002	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP944004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP945003	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP945041	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP945072	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP946010	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP946025	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP946033	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
IACSP952213	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP952288	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP953018	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP953028	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP953104	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP953264	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP955000	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP955011	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP955048	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP955050	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP955110	5	0	3	3	4	0	0	4	2	0	4	0	0	0	3	3	2	4	0	4	41
IACSP956087	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP956100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP956114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22
IACSP966021	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
IACSP966026	0	3	0	0	0	4	4	0	0	4	0	3	4	4	0	0	0	0	3	0	53
PAV9409	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
PO861107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
RB72454	5	3	3	3	4	4	4	4	2	4	4	3	4	4	3	3	2	4	3	4	127
RB825336	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	17
RB835054	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	8
RB835089	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
RB835486	5	3	3	3	4	4	4	4	2	4	4	3	4	4	3	3	2	4	3	4	127
RB855156	0	3	3	0	0	4	4	0	0	4	0	3	4	4	3	0	0	4	3	4	65
RB855453	5	0	3	3	4	0	4	4	0	4	0	4	0	4	3	3	2	4	0	4	65
RB855536	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	7
RB867515	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	9
RB925345	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
SP775181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
SP791011	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9
SP801816	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
SP801842	0	0	0	0	0	0	0	4	0	4	0	0	0	0	0	0	0	0	0	0	72
SP813250	0	0	0	0	0	0	0	4	0	4	4	0	0	0	0	0	0	0	0	0	38
SP832847	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	6
SP835073	5	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16
SP841431	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
SP86155	5	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
SP891115	0	0	3	0	0	0	0	0	0	0	4	3	4	0	3	0	0	4	3	4	28
SP911049	0	0	3	0	4	0	0	0	2	0	0	3	4	0	3	0	0	4	3	4	30

3.1.2 Material Vegetal

Conforme já descrito, a rede experimental analisada foi constituída por 74 genótipos, sendo que nem todos os indivíduos estão presentes em todos experimentos. A frequência de um genótipo em experimentos é variável: há genótipos que foram avaliados mais vezes, os quais foram submentidos a um maior número de ambientes (diferentes locais e várias colheitas), en-

quanto outros foram menos. Como consequência, possuem diferentes número de observações (n).

O conjunto de genótipos analisado na presente tese é composto por clones e variedades de importância comercial. Na Tabela 4 encontram-se todos os genótipos, com destaque para algumas variedades mais cultivadas nas lavouras comerciais brasileiras: RB867515, SP813250, RB855453, RB855156, SP832847, SP801842, RB855536, RB835054, SP801816, IACSP955000 e SP911049. Há também variedades comerciais liberadas pelo IAC: IAC822045, IAC823092, IAC862210, IAC873396, IAC911099, IACSP932060, IACSP933046, IACSP942101, IACSP942094, IACSP944004 e IACSP953028, bem como principais genitores utilizados em cruzamentos e clones promissores do programa de melhoramento do IAC.

Tabela 4 – Conjunto de 74 genótipos considerados nas análises

Genótipos				
IAC811032	IAC911206	IACSP945041	IACSP955110	RB855536
IAC822045	IACSP931020	IACSP945072	IACSP956087	RB867515
IAC823092	IACSP932060	IACSP946010	IACSP956100	RB925345
IAC831313	IACSP933046	IACSP946025	IACSP956114	SP775181
IAC832285	IACSP933050	IACSP946033	IACSP966021	SP791011
IAC832396	IACSP936035	IACSP952213	IACSP966026	SP801816
IAC832405	IACSP942094	IACSP952288	PAV9409	SP801842
IAC834107	IACSP942101	IACSP953018	PO861107	SP813250
IAC853229	IACSP942111	IACSP953028	RB72454	SP832847
IAC862210	IACSP942180	IACSP953104	RB825336	SP835073
IAC873184	IACSP943581	IACSP953264	RB835054	SP841431
IAC873187	IACSP943591	IACSP955000	RB835089	SP86155
IAC873396	IACSP944002	IACSP955011	RB835486	SP891115
IAC911099	IACSP944004	IACSP955048	RB855156	SP911049
IAC911121	IACSP945003	IACSP955050	RB855453	

3.1.3 Caráter Quantitativo

O modelo foi ajustado para o caráter TPH (toneladas de pol por hectare). É uma medida importante para o melhoramento, que mensura conjuntamente a quantidade de sacarose produzida por uma tonelada de cana colhida em um hectare. Envolve, portanto, a produtividade de cana e a concentração de sacarose. Foi estimado pela seguinte expressão (CONSELHO DOS PRODUTORES DE CANA-DE-AÇÚCAR, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO - CONSECANA, 2006):

$$TPH = \frac{PCC \times TCH}{100},$$

em que PCC é o Pol%Cana (porcentagem de sacarose existente na cana) e TCH (tonelada de

cana por hectare). O Pol% de cana foi obtido por:

$$PCC = [Pol\%caldo \times (1 - 0,01 \times Fibra) \times C],$$

em que o Pol%caldo é porcentagem de sacarose contida numa solução de açúcares, e C é um coeficiente de transformação do Pol%caldo em Pol%cana, dado por $C = 1,0313 - 0,00575 \times Fibra$. Fibra da cana foi calculada por $Fibra = 0,08 \times PBU + 0,876$, em que PBU é o peso do bagaço úmido da prensa, em gramas.

3.2 Métodos

3.2.1 Matriz de Parentesco

Com base em trabalhos anteriores (LIMA et al., 2002) e em banco de dados próprio, foi feito um levantamento da genealogia de todos os indivíduos do presente trabalho.

Tal banco de dados contém 1093 indivíduos, composta por clones, genitores, variedades importância comercial e genótipos ancestrais. Todas as relações de parentescos foram incluídas. Para isso, acrescentou-se sistematicamente a informação disponível de todas gerações da cana-de-açúcar até os últimos ancestrais comuns. Partindo dessa genealogia foi construída uma matriz de parentesco de dimensão 1093×1093 . Para tanto, admitiu-se que os genótipos ancestrais não são relacionados e cada genitor contribuiu com igual proporção de alelos para sua progênie.

Foi calculado o coeficiente de parentesco ou coancestria (Φ) entre os 1093 pares de genótipos. A estimativa de Φ , que mensura as relações genéticas entre dois genótipos baseada na análise da genealogia, foi computada calculando-se a probabilidade de dois indivíduos serem idênticos por descendência, utilizando o pacote *Kinship2* do *software* livre *R*. A estimação de Φ é baseada no algoritmo recursivo descrito por Lange (1997), assumindo 0 (zero) para indivíduos não relacionados, 0,50 para o parentesco do indivíduo com ele mesmo e 0,25 para relação pai-filho.

O cálculo de Φ pode ser demonstrado por um exemplo hipotético admitindo dois genitores diplóides e não endogâmico: i e i' com genótipos ab e cd , respectivamente, e com alelos que segregam de forma independente. De uma progênie oriunda do cruzamento entre i e i' os seguintes genótipos são possíveis: ac , ad , bc ou bd , com proporção de 1/4 para cada. Assim, de acordo com Malécot (1969) e Kempthorne (1969):

$$\Phi_{ii'} = \frac{1}{4}[p(a \cong c) + p(a \cong d) + p(b \cong c) + p(b \cong d)],$$

em que $p(a \cong c)$ é a probabilidade dos alelos a e c serem idênticos por descendência.

Após obtida a matriz de parentesco genético da cana-de-açúcar, em seguida desta foi extraída a submatriz simétrica $\mathbf{A}$ de dimensão 74×74 . Essa matriz contém as relação de parentesco entre os genótipos avaliados na tese (Tabela 4). A combinação dois a dois dos 74 genótipos resultou em 2.775 elementos ou coeficientes. Fora da diagonal da matriz $\mathbf{A}$, encontram-se 2.701 coeficientes de parentescos computados entre os pares de indivíduos ($\Phi_{ii'}$), e na diagonal estão os 74 coeficientes de parentesco do indivíduo com ele mesmo (Φ_{ii}), os quais foram anteriormente obtidos pela expressão:

$$\Phi_{ii} = \frac{1}{2}(1 - F_i),$$

em que F_i é o coeficiente de endogamia do indivíduo i .

3.2.2 Modelo Misto

As análises dos experimentos foram realizadas utilizando-se o *software* GenStat 14^o (PAYNE et al., 2009). A princípio buscou-se ajustar o modelo misto adequado que represente realisticamente o padrão natural de resposta dos dados. Isso foi feito comparando-se diferentes estruturas de VCOV para os efeitos genéticos de locais e colheitas (Tabela 5). O modelo matemático descrito aqui foi uma adaptação do modelo usado por Pastina (2010) para análise de MET, de modo que sua descrição seguiu a mesma notação. A notação foi originalmente descrita por Eckermann et al. (2001), Verbyla et al. (2003) e Boer et al. (2007). O modelo estatístico, no qual o sublinhado identifica a variável aleatória, pode ser escrito como:

$$\underline{y}_{ijk r} = \mu + L_j + \underline{C}_k + \underline{LC}_{jk} + \underline{G}_{ijk} + \underline{\varepsilon}_{ijk r} \quad (6)$$

em que $\underline{y}_{ijk r}$ é a resposta fenotípica do i -ésimo genótipo na r -ésima repetição do j -ésimo local e k -ésima colheita; μ é a média geral; L_j é o efeito de local; $\underline{C}_k$ o efeito de colheita; $\underline{LC}_{jk}$ é o efeito da interação local e colheita; $\underline{G}_{ijk}$ é o efeito genético do genótipo i no local j e colheita k ; e $\underline{\varepsilon}_{ijk r}$ os efeitos não genético. Nesse estudo, todos os indivíduos (clones e variedades) foram designados como genótipos, devido ao fato das testemunhas serem diferentes entre os experimentos.

O efeito genético aleatório $\underline{G}_{ijk}$ foi aqui representado por uma notação em minúscula, $\underline{g}_{ijk}$, com vetor $\underline{\mathbf{g}} = (g_{111}, \dots, g_{IJK})$ que possui distribuição normal multivariada com média zero e matriz de VCOV $\mathbf{G}$, escrito na forma de $\underline{\mathbf{g}} \sim N(0, \mathbf{G})$. E o termo $\underline{\varepsilon}_{ijk r}$, foi modelado como:

$$\varepsilon_{ijk r} = b_{jkr} + \eta_{ijk r} \quad (7)$$

em que b_{jkr} é o efeito do bloco r no local j e colheita k ; $\eta_{ijk r}$ é o erro experimental, assumido que $\eta_{ijk r} \sim N(0, \mathbf{R})$, com $\mathbf{R} = \sigma^2 \mathbf{I}$, considerada assim para todos os modelos avaliados.

Todas as interação duplas e triplas entre os efeitos foram também incluídas no modelo, mas por questão de simplificação, não foram explicitadas aqui.

A rotina de análise foi procedida em duas etapas. Na primeira foram ajustados 20 modelos (1 - 20) utilizando a metodologia de modelos mistos descrita por Henderson (1986) e McCulloch e Searle (2001) (Tabela 5). Nesta, cada modelo foi avaliado utilizando-se simultaneamente duas diferentes estruturas de VCOV genética, sendo uma para local ($\mathbf{G}_{J \times J}^L$) e outra para colheita ($\mathbf{G}_{K \times K}^C$). Assumiu-se ausência de parentesco genético entre os indivíduos, dessa forma, a relação entre os genótipos foi modelada por uma matriz identidade ($\mathbf{I}_{I \times I}^G$). As predições do valor genotípico média de TPH (VG) para genótipo teve matriz de VCOV genética dada por $\mathbf{G}_{I \times I} = \mathbf{I}_{I \times I} \sigma_g^2$, em que σ_g^2 é o componente de variância genética. Já na segunda etapa foram ajustados mais 20 modelos (21 - 40). Nestes, foi incluída a informação de parentesco através da matriz $\mathbf{A}$, já comentada no tópico anterior. Na estimação de $\mathbf{A}$ foi assumido a pressuposição de genitores não endogâmicos, e também os efeitos de seleção, mutação e deriva genética foram ignorados (PIEPHO et al., 2008). Assim, a predição do valor genético médio (VA) de TPH para cada genótipo, teve matriz de VCOV genética dada por $\mathbf{G}_{I \times I} = \mathbf{A} \sigma_a^2$, em que σ_a^2 é o componente de variância aditiva (FALCONER e MACKAY, 1996; LYNCH; WALSH, 1998; PIEPHO et al., 2008).

No modelo misto com efeitos aleatórios para o produto cruzado de genótipo e ambientes, a matriz ($\mathbf{G}$) contém as variâncias genéticas para ambientes individuais na diagonal e as covariâncias genéticas entre pares de ambientes fora da diagonal (MARGARIDO, 2011). Diversas formas de estruturação dessa matriz encontram-se na Tabela 1. Smith et al. (2007) trabalhando com dados do tipo MET de cana-de-açúcar, ressaltaram a existência de dois fatores distintos de variação: locais e colheitas, por se tratar de uma cultura perene, e portanto, haver medidas repetidas ao longo do tempo para os mesmos genótipos. Os autores propuseram uma forma alternativa de modelar a matriz ($\mathbf{G}$), através de uma análise combinada de todos os locais e colheitas. Ao invés de fazer uso de uma única matriz de VCOV para combinação fatorial de locais e colheitas, em que cada combinação local $\times$ colheita corresponde a um ambiente, os autores modelaram uma matriz de variâncias e covariâncias para local (denotada aqui como $\mathbf{G}_{J \times J}^L$) e outra para colheita ($\mathbf{G}_{K \times K}^C$), de modo que a matriz $\mathbf{G}$ global pode ser obtida por $\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C$ em que

$\otimes$ denota o produto de Kronecker. Dessa forma, cada uma dessas duas matrizes componentes pode assumir uma das estruturas apresentadas na Tabela 1, e o modelo final pode ter número de parâmetros sensivelmente menor que o obtido quando a matriz ($\mathbf{G}$) não é “fatorada”. Margarido (2011) exemplifica considerando um cenário hipotético em que os genótipos foram avaliados em 2 colheitas e 2 locais. Uma forma não estruturada da matriz ($\mathbf{G}$) com todas as combinações de locais e colheitas (4 ambientes, e) pode ser escrita como:

$$\mathbf{G} = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} & \sigma_{e_1 e_3} & \sigma_{e_1 e_4} \\ \sigma_{e_2 e_1} & \sigma_{e_2}^2 & \sigma_{e_2 e_3} & \sigma_{e_2 e_4} \\ \sigma_{e_3 e_1} & \sigma_{e_3 e_2} & \sigma_{e_3}^2 & \sigma_{e_3 e_4} \\ \sigma_{e_4 e_1} & \sigma_{e_4 e_2} & \sigma_{e_4 e_3} & \sigma_{e_4}^2 \end{bmatrix}.$$

Como tal matriz é simétrica, ou seja, $\sigma_{ij} = \sigma_{ji}$, $\forall i, j \in \{e_1, e_2, e_3, e_4\}$, tem 10 parâmetros no total. Já quando modelada de forma independente, uma para locais e outra para colheitas, têm-se:

$$\mathbf{G}^L = \begin{bmatrix} \sigma_{l_1}^2 & \sigma_{l_1 l_2} \\ \sigma_{l_2 l_1} & \sigma_{l_2}^2 \end{bmatrix} \text{ e } \mathbf{G}^C = \begin{bmatrix} \sigma_{c_1}^2 & \sigma_{c_1 c_2} \\ \sigma_{c_2 c_1} & \sigma_{c_2}^2 \end{bmatrix}$$

Há 3 parâmetros em cada matriz, de modo que há 6 no total. A redução do número de parâmetros torna-se mais acentuada à medida que o número de locais e/ou colheitas aumenta. Por exemplo, para 44 locais e 5 cortes, é necessário estimar 24.310 parâmetros quando utiliza uma única matriz com todas as combinações, e apenas 1.005 quando o produto de Kronecker é empregado. Ocorre redução de quase 96%, aumentando a eficiência do processo de convergência.

Adicionalmente, a relação de independência entre os genótipos ($\mathbf{I}_{I \times I}^G$) (omitida na matriz global $\mathbf{G}$ do exemplo) pode ser substituída por $\mathbf{A}$ (SMITH et al., 2007).

Nesse contexto, a matriz de VCOV genética global $\mathbf{G}$ foi modelada através do produto direto entre três matrizes. Sendo duas de variâncias e covariâncias (uma para locais e outra para colheitas) e uma que modela a relação existente entre os genótipos.

Para os 20 primeiros modelos foi assumido ausência de parentesco genético, e a matriz de VCOV $\mathbf{G}$ foi obtida da seguinte forma:

$$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G \quad (8)$$

e, para os modelos com a informação de parentesco (21 a 40):

$$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G \quad (9)$$

em que, $\mathbf{G}^L$, $\mathbf{G}^C$ e $\mathbf{A}^G$ são as matrizes variâncias covariâncias genética para locais, colheitas e de parentesco genético, respectivamente. Essas duas condições (8 e 9) foram verificadas a fim de conhecer a contribuição dada com a inclusão da genealogia. Uma vez que tais modelos são mais realistas, espera-se um maior poder preditivo, e portanto o aumento da eficiência de seleção.

Os 40 modelos foram examinados e comparados através dos critérios AIC e BIC. Segundo esses critérios o modelo que melhor se ajusta aos dados é aquele que tem o menor valor dessas estatísticas. Após selecionado o “melhor” modelo, foram obtidas as predições do valor genotípico médio (VG) e/ou do valor genético médio (VA) de TPH, e, em seguida, foi definido o ranqueamento dos genótipos.

Como a realização de análises de variância conjunta é inviável, dado o elevado grau de desbalanceamento, para efeito de comparação, estimou-se a média aritmética de TPH. Esse tipo de análise, é o procedimento estatístico comumente usado por melhorista de cana para avaliar o potencial produtivo médio dos genótipos nos grupos de experimentos, especialmente nos casos em que nem todos os genótipos são avaliados em todos experimentos. Para o conjunto de dados analisado, a extensão do desbalanceamento tornou-se maior devido existir diferentes genótipos, números de repetições e também diferentes números de colheitas. Retrata-se então, uma situação frequentemente enfrentada pelos programas de melhoramento. Em geral, para contornar esse impasse, uma alternativa que tem sido utilizada, na maioria da vezes, é simplesmente o cálculo da média aritmética simples para cada indivíduo. Entretanto, a média de cada genótipo é calculada com números diferentes de observações devido ao grande desbalanceamento dos dados. Assim, análogo aos programas de melhoramento, foi calculado a média aritmética (tratada aqui como Modelo Tradicional), e, em seguida, foi estabelecido o ranqueamento dos genótipos. A intensidade da associação entre os modelos em função das coincidências de ranqueamento foi determinada calculando-se o coeficiente de correlação de *Spearman* (ρ), adequado por levar em consideração a ordem dos dados e não o seu valor intrínseco. Assim, os valores genéticos preditos obtidos pelos modelos Tradicional, 1 e 11, foram ordenados e comparados. As coincidências de ranqueamento foram verificadas calculando-se ρ entre os pares de modelos.

Teoricamente, espera-se quanto maior o valor de ρ ($\rho \rightarrow 1$) maior a associação, ou seja, maior a coincidência de ranqueamento entre os modelos. Nesse sentido, o menor o valor de ($\rho \rightarrow 0$), indica baixa associação, ou seja, menor a coincidência, os modelos ranqueiam de forma diferente.

Em relação a obtenção das estimativas do modelo, os efeitos fixos foram estimados pelo *Best Linear Unbiased Estimator* (BLUE), e os aleatórios por *Best Linear Unbiased Predictor* (BLUP), já implementados no *software* Genstat. O processo de estimação ocorreu via Máxima Verossimilhança e Máxima Verossimilhança Restrita (REML)(PATTERSON e THOMPSON, 1971).

Tabela 5 – Diferentes modelos analisados para a matriz de variâncias e covariâncias $\mathbf{G}$ do modelo misto

Matrix $\mathbf{G}$	Modelo				n_{PAR} em $\mathbf{G}^\dagger$
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G$	1)	ID	$\otimes$	ID $\otimes$ ID	1
	2)	DIAG	$\otimes$	ID $\otimes$ ID	$(J + 1) - 1$
	3)	FA1	$\otimes$	ID $\otimes$ ID	$(2J + 1) - 1$
	4)	UNST	$\otimes$	ID $\otimes$ ID	$\left[\frac{J(J+1)+2}{2} \right] - 1$
	5)	ID	$\otimes$	DIAG $\otimes$ ID	$(1 + K) - 1$
	6)	DIAG	$\otimes$	DIAG $\otimes$ ID	$(J + K) - 1$
	7)	FA1	$\otimes$	DIAG $\otimes$ ID	$(2J + K) - 1$
	8)	UNST	$\otimes$	DIAG $\otimes$ ID	$\left[\frac{J(J+1)+2K}{2} \right] - 1$
	9)	ID	$\otimes$	AR1 $\otimes$ ID	2
	10)	DIAG	$\otimes$	AR1 $\otimes$ ID	$(J + 2) - 1$
	11)	FA1	$\otimes$	AR1 $\otimes$ ID	$(2J + 2) - 1$
	12)	UNST	$\otimes$	AR1 $\otimes$ ID	$\left[\frac{J(J+1)+4}{2} \right] - 1$
	13)	ID	$\otimes$	AR1 _{Het} $\otimes$ ID	$(K + 2) - 1$
	14)	DIAG	$\otimes$	AR1 _{Het} $\otimes$ ID	$(J + K + 1) - 1$
	15)	FA1	$\otimes$	AR1 _{Het} $\otimes$ ID	$(2J + K + 1) - 1$
	16)	UNST	$\otimes$	AR1 _{Het} $\otimes$ ID	$\left[\frac{J(J+1)+2(K+1)}{2} \right] - 1$
	17)	ID	$\otimes$	UNST $\otimes$ ID	$\left[\frac{2+K(K+1)}{2} \right] - 1$
	18)	DIAG	$\otimes$	UNST $\otimes$ ID	$\left[\frac{2J+K(K+1)}{2} \right] - 1$
	19)	FA1	$\otimes$	UNST $\otimes$ ID	$\left[\frac{4J+K(K+1)}{2} \right] - 1$
	20)	UNST	$\otimes$	UNST $\otimes$ ID	$\left[\frac{J(J+1)+K(K+1)}{2} \right] - 1$
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G$	21)	ID	$\otimes$	ID $\otimes$ Kinship	1
	22)	DIAG	$\otimes$	ID $\otimes$ Kinship	$(J + 1) - 1$
	23)	FA1	$\otimes$	ID $\otimes$ Kinship	$(2J + 1) - 1$
	24)	UNST	$\otimes$	ID $\otimes$ Kinship	$\left[\frac{J(J+1)+2}{2} \right] - 1$
	25)	ID	$\otimes$	DIAG $\otimes$ Kinship	$(1 + K) - 1$
	26)	DIAG	$\otimes$	DIAG $\otimes$ Kinship	$(J + K) - 1$
	27)	FA1	$\otimes$	DIAG $\otimes$ Kinship	$(2J + K) - 1$
	28)	UNST	$\otimes$	DIAG $\otimes$ Kinship	$\left[\frac{J(J+1)+2K}{2} \right] - 1$
	29)	ID	$\otimes$	AR1 $\otimes$ Kinship	2
	30)	DIAG	$\otimes$	AR1 $\otimes$ Kinship	$(J + 2) - 1$
	31)	FA1	$\otimes$	AR1 $\otimes$ Kinship	$(2J + 2) - 1$
	32)	UNST	$\otimes$	AR1 $\otimes$ Kinship	$\left[\frac{J(J+1)+4}{2} \right] - 1$
	33)	ID	$\otimes$	AR1 _{Het} $\otimes$ Kinship	$(K + 2) - 1$
	34)	DIAG	$\otimes$	AR1 _{Het} $\otimes$ Kinship	$(J + K + 1) - 1$
	35)	FA1	$\otimes$	AR1 _{Het} $\otimes$ Kinship	$(2J + K + 1) - 1$
	36)	UNST	$\otimes$	AR1 _{Het} $\otimes$ Kinship	$\left[\frac{J(J+1)+2(K+1)}{2} \right] - 1$
	37)	ID	$\otimes$	UNST $\otimes$ Kinship	$\left[\frac{2+K(K+1)}{2} \right] - 1$
	38)	DIAG	$\otimes$	UNST $\otimes$ Kinship	$\left[\frac{2J+K(K+1)}{2} \right] - 1$
	39)	FA1	$\otimes$	UNST $\otimes$ Kinship	$\left[\frac{4J+K(K+1)}{2} \right] - 1$
	40)	UNST	$\otimes$	UNST $\otimes$ Kinship	$\left[\frac{J(J+1)+K(K+1)}{2} \right] - 1$

Modelos (1-20) utilizam o produto direto de matrizes de variância-covariância para experimento e corte, assumindo ausência de parentesco entre os genótipos, $\mathbf{A}$. Modelos (21-40) utilizam o produto direto de matrizes de variância-covariância para experimento e corte, assumindo parentesco genético entre os genótipos. † : o número de parâmetros para os modelos (1-40), correspondem à soma dos números de parâmetros das matrizes componentes menos o número de restrições para assegurar a identificabilidade do modelo. I é o número de genótipos, J é o número de experimentos (loais) e K o número de cortes.

4 RESULTADOS

4.1 Análise Usando Modelo Fixo

A Tabela 6 mostra os resultados da análise descritiva para variável TPH. Foram obtidas medidas de tendência central como a média aritmética e mediana, e também medidas de dispersão tais como a variância, o desvio padrão e os valores máximos e mínimos. A partir das estimativas da média foi estabelecido o *ranking* dos 74 genótipos. Em geral, a produtividade média de TPH variou 7,689 t.ha⁻¹ a 18,607 t.ha⁻¹, com variância de 4,998 t².ha⁻² a 42,521 t².ha⁻². Os resultados apontaram os seguintes genótipos como os mais produtivos (com suas respectivas médias de variâncias): RB925345 ($\bar{x} = 18,607$ t.ha⁻¹, $\sigma^2 = 28,350$ t².ha⁻²), RB835089 ($\bar{x} = 18,272$ t.ha⁻¹, $\sigma^2 = 42,521$ t².ha⁻²), RB855536 ($\bar{x} = 17,794$ t.ha⁻¹, $\sigma^2 = 10,884$ t².ha⁻²), RB825336 ($\bar{x} = 17,422$ t.ha⁻¹, $\sigma^2 = 11,823$ t².ha⁻²), IAC873184 ($\bar{x} = 16,978$ t.ha⁻¹, $\sigma^2 = 16,067$ t².ha⁻²), IAC822045 ($\bar{x} = 16,964$ t.ha⁻¹, $\sigma^2 = 17,030$ t².ha⁻²), IAC823092 ($\bar{x} = 16,813$ t.ha⁻¹, $\sigma^2 = 14,881$ t².ha⁻²), SP86155 ($\bar{x} = 16,745$ t.ha⁻¹, $\sigma^2 = 21,835$ t².ha⁻²), IAC873396 ($\bar{x} = 16,667$ t.ha⁻¹, $\sigma^2 = 22,437$ t².ha⁻²) e IAC911099 ($\bar{x} = 16,372$ t.ha⁻¹, $\sigma^2 = 22,461$ t².ha⁻²), (Tabela 6 e Figura 2). Desses, RB925345, RB835089, SP86155, IAC873396 e IAC911099 tiveram variância alta, o que leva a pensar em provável inconsistência de suas produtividades médias; portanto, o desempenho observado poderá não ser refletido em condições não experimentais. Outra possibilidade é a elevada presença de interação entre genótipos e ambientes.

Evidentemente a média aritmética é uma estatística interessante para resumir dados estatísticos e também para fornecer uma compreensão maior das informações sobre o fenômeno estudado. Mas é importante salientar que a média é uma medida estatística que possui a desvantagem de ser fortemente afetada por valores extremos. A dispersão dos valores de TPH com seus valores máximos e mínimos para os 74 genótipos pode ser também verificada na Figura 2. Há genótipo que apresenta valor TPH extremo, muito alto ou muito baixo, como por exemplo os genótipos IAC823092, IAC911099, RB72454, PO861107, IACSP966021, IACSP943591, RB855453, IACSP953028, IACSP953104 e IACSP942180 (representados por box plots vermelho) com valores superiores a 29 t.ha⁻¹; os genótipos IACSP933050, SP911049, IACSP966026, IACSP952288, IACSP945072, IACSP936035 e PAV9409 (os box plots verde), com valores inferiores a 3 t.ha⁻¹; e ainda os genótipos IACSP956087, IACSP943581, IAC911121, IAC911206 (os box plots verde e vermelho) com valores inferior 3 t.ha⁻¹ e superior a 29 t.ha⁻¹. Nos três casos os extremos certamente influenciaram produtividade média dos genótipos.

A presença de valores de TPH discrepantes é um dos aspectos a ser considerado. As observações com presença de *outliers* altas e baixas estão destacadas com as cores laranja e azul escuro, respectivamente, nas duas barras à esquerda da Figura 2. Nota-se que cerca de 51% dos genótipos tem observações que são *outliers* altos e/ou baixos, ou seja, alguns genótipos tem valores de TPH fora do padrão; são os pontos projetados afastados do conjunto de observações de cada genótipo. Teoricamente, a existência de algum valor extremo muito alto tende a aumentar a média, enquanto valores muito baixo tende a diminuir-la. Assim, é possível que a média de TPH de alguns genótipos possa estar superestimada ou subestimada.

Outro aspecto a ser considerado é os diferentes números de observações por genótipos que ocorreu devido o desbalanceamento dos dados. Para fins de comparação, a média deveria ser estimada com o mesmo número de observações para todos genótipos, nos mesmos ambientes. O número de observações (N) por genótipo variou de 12 a 471 (com $N = N_r \times N_c \times N_l$, em que N_r , N_c e N_l , são o número de repetições, de colheitas e locais, respectivamente). Essa variação é bastante grande e o menor número de observações pode comprometer a precisão da estimativa da média. É possível por exemplo que a média de produtividade dos genótipos RB925345, RB835089, RB855536, SP86155, IACSP944004, IACSP933046, IACSP933050, RB867515, SP835073, SP841431, RB835054, SP891115, IACSP946033 e IACSP942094 possa ser imprecisa, não refletindo portanto o “verdadeiro” potencial produtivo dos indivíduos. Entre os 10 “melhores” genótipos, nota-se que RB835089, classificada como o segundo melhor genótipo, teve apenas 12 observações ($N_r = 6$, $N_c = 2$ e $N_l = 1$). Suas medidas possuem variância altíssima (42,521 t.ha⁻¹), com mediana de 16,795 t.ha⁻¹. O baixo número de observações aliado a alta variação indica que provavelmente sua produtividade média seja inconsistente. Outro exemplo é RB925345, com 18 observações ($N_r = 3$, $N_c = 3$ e $N_l = 2$), que foi eleito o genótipo mais produtivo, mas possui variância alta. Apesar da mediana indicar que 50% de sua produtividade está acima de 17,810 t.ha⁻¹, sua produtividade média pode ser irrealista.

Número diferentes de colheitas também pode afetar a estimativa da média. O perfil produtivo médio de um genótipo varia em função da natureza da colheita (planta, soca e ressoca) (FREITAS, 2007). Diferentes números de observações tomados na planta, soca ou ressoca pode gerar médias gerais inconsistentes.

Portanto, quando se tem medidas discrepantes e/ou diferente números de observações, a classificação estabelecida pelo ranqueamento dos genótipos em função da média pode não ser verdadeira. A média por si só, não é uma estatística adequada para avaliar o desempenho dos genótipos. Seu emprego para comparação e ranqueamento dos genótipos deve ser visto com cautela. Adicionalmente, a média fenotípica fornecida pelo modelo tradicional é uma estimativa

bastante limitada, não servindo para realização predições futuras. Para tanto, faz-se necessário a aplicação de metodologia mais adequada para análise de grupos de experimentos de cana, e o uso de modelo linear misto, certamente é uma alternativa mais apropriada e deve ser considerada.

4.2 Coeficiente de Parentesco

Nesse tópico serão sintetizados os resultados referentes às relações de parentesco, detalhando principalmente, a contribuição dada por cada indivíduo em termos de parentesco genético. A ideia é mostrar as relações de parentesco existentes. O entendimento da base genética, bem como das relações entre os genótipos de cana-de-açúcar, é importante para o programa de melhoramento, especialmente para escolha de genitores que participarão dos futuros cruzamentos. De alguma forma, espera-se que essa informação seja relevante e possa contribuir para orientação e direcionamento de novos cruzamentos.

A matriz **A** resultante possui grandes dimensões (74×74), sua representação numérica é de difícil visualização, nessa circunstância, optou-se por representá-la de forma gráfica (Figura 3).

Observa-se um gradiente de cores que está associado as diferentes relações de parentesco. Para efeito de interpretação, o azul claro indica que o parentesco entre os genótipos não existe ou é muito baixo ($\Phi_{ii'} \rightarrow 0$). E, a medida que as cores tornam-se mais próxima do vermelho, paralelamente a relação torna-se gradativamente maior ($\Phi_{ii'} \rightarrow 1$). De forma geral, nota-se que os genótipos pouco ou não aparentados foram destacados por azul claro, enquanto os mais aparentados foram destacados por vermelho. Evidentemente, o coeficiente de parentesco de um indivíduo com ele mesmo tende a ser maior do que o coeficiente entre indivíduos. Por essa razão, a diagonal da matriz assumiu em geral, diferentes tonalidades de verdes.

Para melhor entendimento, é interessante avaliar conjuntamente as Figuras 3 e 4. A Figura 4 é um dendrograma dos 74 genótipos construído a partir de informação de dissimilaridade genética, calculada por $d = 1 - \Phi$. Quando $d = 1$, a dissimilaridade é máxima, ou seja, não há parentesco entre os indivíduos, e, a medida que a dissimilaridade diminui ($d \rightarrow 0$), o parentesco torna-se maior. Isto posto, é possível verificar se um dado genótipo possui parentesco com outros indivíduos e identificar quais são os seus parentes (Figura 3). Reunida essa informação, é possível destacar todos os indivíduos relacionados e definir grupos em função do grau de relacionamento (Figura 4).

Para interpretação dos resultados, buscou-se seguir a ordem dos genótipos disposta na Figura 4. Estes foram detalhados no sentido de baixo para cima, e eventualmente, para esclarecer alguns parentescos, foram citados genótipos que não pertence ao conjunto dos 74 avaliados, tais

Figura 2 – Box plots dos dados de TPH para 74 genótipos

Tabela 6 – Análise descritiva de TPH para 74 genótipos com os correspondentes número de observações, média, variância, desvio padrão, mediana, mínimo, máximo e o *Ranking* (Continua)

Genótipos	Nº de observações	Média de TPH	Variância	Desvio padrão	Mediana	Mínimo	Máximo	Ranking
RB925345	18	18,607	28,350	5,325	17,810	10,270	28,500	1^o
RB835089	12	18,272	42,521	6,521	16,795	9,880	28,600	2^o
RB855536	36	17,794	10,884	3,299	18,210	12,300	22,620	3^o
RB825336	93	17,422	11,823	3,438	17,750	9,930	24,010	4^o
IAC873184	180	16,978	16,067	4,008	16,655	7,180	27,880	5^o
IAC822045	180	16,964	17,030	4,127	16,945	6,800	28,790	6^o
IAC823092	180	16,813	14,881	3,858	16,865	5,570	31,160	7^o
SP86155	60	16,745	21,835	4,673	17,530	7,830	28,880	8^o
IAC873396	222	16,667	22,437	4,737	16,325	5,730	29,310	9^o
IAC911099	159	16,372	22,461	4,739	16,160	3,820	34,000	10^o
IAC873187	180	16,101	11,919	3,452	15,550	7,930	25,850	11 ^o
IACSP944004	66	15,900	20,102	4,484	16,010	6,290	26,190	12 ^o
SP801842	305	15,688	15,176	3,896	15,970	4,860	26,900	13 ^o
IAC832285	180	15,586	13,020	3,608	15,430	7,080	27,460	14 ^o
IACSP933046	66	15,410	10,848	3,294	15,040	6,080	22,240	15 ^o
RB72454	471	15,409	24,888	4,989	15,610	3,610	35,280	16 ^o
IACSP933050	66	15,356	27,413	5,236	15,760	2,020	28,650	17 ^o
IAC811032	180	14,937	15,191	3,898	14,765	5,930	27,600	18 ^o
RB867515	27	14,785	9,597	3,098	14,700	7,380	20,130	19 ^o
SP835073	48	14,762	9,525	3,086	15,160	8,510	22,580	20 ^o
SP841431	24	14,503	11,014	3,319	13,675	9,480	21,250	21 ^o
RB835054	24	14,467	24,343	4,934	14,105	6,650	24,380	22 ^o
SP891115	84	14,465	22,682	4,763	14,460	3,300	23,510	23 ^o
IACSP946033	12	14,446	4,998	2,236	14,185	11,700	17,830	24 ^o
IAC862210	231	14,444	18,199	4,266	13,900	6,160	26,970	25 ^o
IACSP942094	66	14,273	11,813	3,437	14,250	4,980	20,190	26 ^o
IACSP955000	123	14,270	19,698	4,438	14,170	5,890	24,920	27 ^o
PO861107	180	14,193	15,125	3,889	13,785	5,610	32,430	28 ^o
IACSP966021	159	14,192	21,698	4,658	13,400	5,020	33,010	29 ^o
IACSP932060	159	14,003	17,029	4,127	13,230	6,010	29,550	30 ^o
IACSP956087	66	13,959	25,892	5,088	14,080	2,610	30,160	31 ^o
SP813250	114	13,934	15,955	3,994	14,195	4,860	22,800	32 ^o
IAC831313	180	13,926	16,971	4,120	13,675	5,680	28,540	33 ^o
IACSP955011	123	13,925	16,099	4,012	13,960	5,970	24,260	34 ^o
IACSP946025	66	13,892	20,242	4,499	13,630	5,190	25,620	35 ^o
SP775181	32	13,833	9,067	3,011	13,405	8,530	20,590	36 ^o
IAC834107	179	13,794	17,405	4,172	13,250	5,120	28,190	37 ^o
IACSP942101	66	13,714	20,520	4,530	13,555	3,460	22,830	38 ^o
IACSP943591	159	13,688	25,429	5,043	13,610	4,890	37,380	39 ^o
IAC832396	178	13,631	14,001	3,742	13,335	3,850	23,120	40 ^o
SP791011	54	13,327	8,149	2,855	13,100	7,400	21,740	41 ^o
IAC832405	180	13,278	10,876	3,298	13,235	6,740	22,280	42 ^o
RB855453	195	13,229	22,556	4,749	12,850	4,010	38,180	43 ^o
IACSP943581	159	13,158	27,700	5,263	12,360	2,230	40,440	44 ^o
IACSP953028	159	13,146	17,398	4,171	12,590	6,050	33,650	45 ^o
IACSP931020	159	13,117	17,930	4,234	13,230	3,620	25,830	46 ^o
IACSP953018	123	13,057	15,377	3,921	12,810	4,160	21,040	47 ^o
RB835486	429	12,983	17,267	4,155	12,480	4,400	29,190	48 ^o
IACSP945041	66	12,977	11,227	3,351	13,395	3,460	19,660	49 ^o
RB855156	195	12,844	9,791	3,129	12,780	4,670	20,400	50 ^o
SP911049	90	12,741	15,619	3,952	12,475	2,730	23,040	51 ^o
IACSP944002	66	12,704	18,325	4,281	12,195	3,260	28,600	52 ^o
IACSP953264	123	12,666	17,235	4,152	12,070	4,930	23,730	53 ^o
IAC853229	178	12,604	13,446	3,667	12,755	3,950	22,330	54 ^o
IACSP953104	159	12,530	15,107	3,887	12,190	6,280	31,370	55 ^o
IACSP952213	123	12,509	19,506	4,417	12,170	4,230	23,350	56 ^o
IACSP942111	159	12,423	15,573	3,946	11,670	5,390	27,810	57 ^o
IACSP955050	123	12,261	16,747	4,092	12,050	5,010	23,700	58 ^o
IACSP955048	123	12,030	11,752	3,428	12,200	5,230	19,290	59 ^o
IACSP955110	123	11,926	12,897	3,591	11,410	3,230	22,700	60 ^o
SP832847	18	11,770	15,611	3,951	11,130	5,270	17,730	61 ^o
IACSP942180	159	11,726	18,018	4,245	11,270	3,700	36,670	62 ^o
IACSP966026	159	11,703	23,587	4,857	11,620	2,270	23,140	63 ^o

Tabela 6 – Análise descritiva de TPH para 74 genótipos com os correspondentes número de observações, média, variância, desvio padrão, mediana, mínimo, máximo e o *Ranking* (Conclusão)

Genótipos	Nº de observações	Média de TPH	Variância	Desvio padrão	Mediana	Mínimo	Máximo	Ranking
IACSP956100	66	11,620	14,978	3,870	11,935	4,350	26,590	64 °
IAC911121	159	11,568	18,002	4,243	11,410	2,830	34,140	65 °
IACSP956114	66	11,469	9,734	3,120	11,025	5,520	23,720	66 °
IACSP952288	123	11,395	13,388	3,659	11,730	2,060	18,190	67 °
IAC911206	159	11,212	19,102	4,371	10,400	2,520	32,050	68 °
IACSP945003	66	10,294	5,667	2,381	10,210	4,130	16,630	69 °
IACSP946010	66	10,253	7,618	2,760	10,205	4,410	16,110	70 °
SP801816	21	9,687	8,997	2,999	10,360	3,520	12,980	71 °
IACSP945072	66	9,386	11,578	3,403	9,420	1,530	18,250	72 °
IACSP936035	66	8,568	6,707	2,590	8,140	2,750	14,030	73 °
PAV9409	15	7,689	13,672	3,698	6,360	2,640	13,500	74 °

Os dez melhores genótipos estão destacados em negrito.

genótipos foram diferenciados com um “*” no final.

Verificou-se que RB835054 e RB835089 (que estão próximas na Figura 4) são irmãos completos, ambos são meios-irmãos de RB855156, e as três são filhos de RB72454; IACSP953028 é filho de RB855156, que por sua vez é filho de RB72454, portanto, aparentada com RB835054 e RB835089; IACSP943581, RB855536, IACSP943591 e IACSP953264 são irmãos completos, filhos do cruzamento entre SP701143* e RB72454; RB867515 é filho de RB72454, portanto tem parentesco com IACSP953028 e é meio-irmão dos outros anteriores citados. Até este ponto, nota-se que existe um grupo de 10 genótipos aparentados, isso representando 13, 51% dos genótipos que foram utilizado em experimentação pelo IAC durante um período de doze anos.

Continuando, têm-se IACSP953018, filho de SP842189*, e meio-irmão de IACSP953028; IACSP956087 é filho de IAC873396. IAC873184, IAC834107, IAC873187 e SP832847, junto com IAC873396 e com o quarteto formado por IACSP943581, RB855536, IACSP943591 e IACSP953264 são irmãos, filhos de SP701143*. Desse grupo, os três primeiros são irmãos completos entre si e meios-irmãos de IAC823092, filhos de IAC6812*, enquanto os seis seguintes são meios-irmãos; IACSP956114 é filho de IAC873187, portanto, é parente do grupo dos nove anteriormente citados; IACSP955048, IACSP953104 e IACSP955050 são irmãos completos, filhos do cruzamento entre SP823697* e IAC873184, já citado, portanto, possui parentesco com o grupo dos nove e também com IACSP956114, já citado.

Em seguida tem-se IACSP931020, IACSP933046 e IACSP933050, todos meios-irmãos, filhos de SP791011; IAC832396 é meio-irmão de SP791011, portanto, tem parentesco com os três últimos citados anteriormente; enquanto IAC822045 e IAC811032 possuem pais que são irmãos completos; IAC862210 tem parentesco com SP791011, portanto, com todos os genótipos a ela relacionado, e também com IAC811032.

Na sequência, tem-se IACSP945041, IACSP942111, IACSP946010 e IACSP952288, meios-irmãos, filhos de SP803280*, sendo que IACSP952288 é filho também de RB835486; IACSP955011 e IACSP936035 são meios-irmãos, filhos de SP801842 que é irmão completo da SP801816, os dois últimos são meios-irmãos de SP835073, filhos de SP711088*; RB855453 possuem parentesco distante ($d \rightarrow 1$) com IAC832285 e com vários outros.

A partir desse ponto, o parentesco é muito baixo, assim, visualizar a Figura 3 torna-se mais interessante. Verifica-se que a intensidade de azul claro é maior e a distinção não é tão clara, dificultando detectar a existência de parentesco. Mas embora as relações de parentesco sejam baixas, elas existem, e ainda é possível visualizá-las. Similarmente, IAC832285 e IAC832405 tem parentesco distante com SP791011 e muitos outros, as relações existentes também podem ser visualizadas (Figura 3). Em seguida tem-se IACSP955110 e IACSP952213 que são irmãos completos e estão lado a lado no dendrograma; IACSP946033 e SP911049 que são meios-irmãos, ambos filhos de SP813250; IACSP956100 possui parentesco com outros, porém extremamente baixo ($\Phi_{ii'} \rightarrow 0$), quase imperceptível (Figura 3), o mesmo se aplica a RB825336. Em seguida tem-se IACSP966021 e IACSP942094 que são meios-irmãos, filhos de SP847017*; IACSP932060 que possui parentesco extremamente baixo com outros, praticamente imperceptível (Figura 3). Já IACSP966026 e IACSP955000 são meios-irmãos, filhos de SP80185*; o primeiro é meio-irmão de IACSP944004 e IACSP944002, juntos são filhos de SP826108*. IACSP944004 e IACSP944002 são irmãos completos, sendo meios-irmãos de IACSP942101, IACSP942180 e IACSP945003, todos filhos de SP775181; IAC911099 é meio-irmão da IACSP942101 e filhos de RB785148*; IACSP946025 é meio-irmão de IACSP942180, seu único parente, ambos filhos de SP842029*.

E por fim, tem-se os indivíduos IAC831313, SP841431, IAC853229, PO861107, SP86155, SP891115, IAC911121, IAC911206, RB925345, IACSP945072 e PAV9409 que tiveram coeficiente de parentesco igual a zero ($\Phi_{ii'} = 0$), com distância máxima dos demais ($d = 1$) (Figura 4). Houve a formação de um grupo isolado (Figura 4), numa faixa completamente azul claro (Figura 3). Essa disposição sugere que, para genealogia em questão, não houve evidência parentesco.

Os valores obtidos do coeficiente de parentesco entre os 74 genótipos tiveram uma média de 0,0293, com valores variando de 0 a 0,35. E a maior parte dos valores concentrou-se no intervalo de 0 à 0,05 (Figura 5). Lima et al., 2002 avaliaram um conjunto com 83 genótipos de cana-de-açúcar, e encontraram valores do coeficiente de parentesco que variando de 0 à 0.503, com uma média de 0.057.

Os genótipos RB835054 e RB835089 ambos mostraram o maior parentesco médio com os

demais genótipos ($\Phi_{ii'} = 0.0716$), seguidos de SP791011 ($\Phi_{ii'} = 0.0701$), indicando que no período de 12 anos os três foram bastante utilizados no programa de melhoramento IAC (Figura 6). Verifica-se ainda que 44 genótipos (59,4%) possuem uso moderado ($0.02 < \Phi_{ii'} < 0.05$) e 27 genótipos (36,5%) com baixa utilização ($\Phi_{ii'} < 0.02$) (Figura 6 e 7).

4.3 Seleção do Modelo

Nesse tópico utilizando a abordagem de modelos misto, buscou-se ajustar um modelo alternativo apropriado para análise de grupos de experimentos de cana-de-açúcar, incluindo a informação de parentesco. Trata-se de um modelo mais realista, que inclui todas as informações de covariância genética existente entre indivíduos, colheitas e locais, que normalmente são ignoradas pela abordagem tradicional de análise de variância (modelo fixo). A configuração teórica da metodologia de modelos misto é apropriada para análise de dados experimentais desbalanceados, independentemente da natureza do desbalanceamento, seja por falta, perda de parcelas ou por delineamento não ortogonal. As particularidades da abordagem inerente a modelagem de informação de VCOV genética para local e colheita confere a elaboração de modelos de alta capacidade preditiva. Quando estruturas de VCOV para local e corte são selecionadas adequadamente, espera-se que o modelo resultante forneça previsões mais acuradas dos valores genotípicos. Tais valores poderão ser usados para fazer inferências sobre o comportamento produtivo atual e futuro dos indivíduos. Portanto, o uso de modelos mais realistas, permitirá estabelecer o ranqueamento dos indivíduos, aumentando assim, a eficiência de seleção.

No contexto dos modelos mistos, é possível inserir ainda no modelo a informação de covariância genética aditiva entre os genótipos. Essa informação é incorporada via matriz de parentesco genético, $\mathbf{A}_{I \times I}^G$, já comentada no tópico anterior. O modelo com a inclusão de informação de parentesco combinada com a modelagem da matriz de VCOV genética para local e corte torna-se ainda mais realista. É importante mencionar que a inclusão de informação de parentesco genético computa informação de variância genética aditiva e, portanto, permite previsões do valor genético. Essa medida pode ser usada para identificar os melhores genitores de cana-de-açúcar e definir os futuros cruzamentos.

A Tabela 7 mostra os resultados das análises de 40 modelos diferentes, considerando a matriz $\mathbf{G}$ de VCOV genética. A matriz $\mathbf{G}$ está configurada de duas formas: na primeira, ela é composta pelo produto direto entre três matrizes, sendo duas matrizes de VCOV genética para locais e colheitas, e a terceira é a matriz identidade para genótipos, que assume que não existe parentesco entre eles ($\mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G$). Na segunda configuração as duas primeiras matrizes que

Figura 3 – Heatmap construído a partir da matriz de parentesco genético (A) dos 74 genótipos

Figura 4 – Dendrograma dos 74 genótipos de cana-de-açúcar, obtido calculando-se ($d = 1 - \Phi$) como medida de distância genética e utilizando o método de agrupamento UPGMA

Figura 5 – Distribuição do coeficiente de parentesco entre os genótipos de cana-de-açúcar

compõem a matriz $\mathbf{G}$ são as mesma que anterior, porém a terceira é a matriz de parentesco genético, que considera a relação de parentesco entre os indivíduos ($\mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G$). Para cada configuração foram testados 20 modelos. Os modelos de 1 a 20 estão de acordo com a primeira configuração ($\mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G$), enquanto os modelos de 21 a 40 estão conforme a segunda configuração ($\mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G$).

Diferentes estruturas de VCOV foram testadas para local e colheita (Tabela 7). Para especificação de cada estrutura, levou-se em consideração o conhecimento prévio do pesquisador sobre os padrões de respostas esperado das observações tomadas nos diferentes locais e colheitas. Assim, foram testadas as estruturas ID, DIAG, FA1 e UNST para a matriz $\mathbf{G}_{J \times J}^L$ de locais, e ID, DIAG, AR1, AR1_{Het} e UNST para matriz $\mathbf{G}_{K \times K}^C$ de colheitas. Num primeiro momento, a matriz de relações entre os indivíduos foi sempre a mesma, $\mathbf{I}_{I \times I}^G$, ignorando o parentesco entre o genótipos, ou a matriz $\mathbf{A}_{I \times I}^G$, admitindo que existe parentesco entre eles.

De acordo com o critério AIC e BIC, em geral os modelos com a inclusão da informação de parentesco (21 - 40), foram melhores quando comparados aos correspondentes (1 - 20) sem esta informação. Esse resultado confirma que a informação de parentesco contribuiu para melhoria do modelo, tornando-o mais realista. Uma explicação é que os genótipos compartilham locos que

Figura 6 – Distribuição de valores do coeficiente de parentesco médio de cada um dos 74 genótipos de cana-de-açúcar com todos os demais do conjunto avaliado

Figura 7 – Distribuição do coeficiente de parentesco médio de cada um dos 74 genótipos de cana-de-açúcar

são idênticos por descendência, e através de cruzamentos, essa herança genética é transmitida para novas gerações. Finalmente há formação de uma grande “rede de relacionamento” com indivíduos de diferentes graus de parentesco. Com essas considerações, modelos que assumem independência entre os genótipos, não considera as relações existentes, não aproveitando portanto, informações que já estão à disposição sem custos adicionais. Em contrapartida, modelos que incluem o parentesco oferecem aos genótipos a oportunidade de empregar a informação da genealogia; mesmo genótipos com pouca informação genealógica são beneficiados.

Adicionalmente, a informação de genealogia no modelo dá uma expectativa, em termos de desempenho produtivo, do valor genético que o indivíduo tem e transmite para sua descendência. Para tanto, os efeitos genéticos aditivo (valor genético) dos genitores transmitido para progênie são incluídos no modelo via matriz de parentesco. Assim, o modelo resultante fornece previsões do valor genético. No melhoramento de cana-de-açúcar, as previsões geradas por esse modelo mensura o quanto de variância genética aditiva um genótipo poderá transmitir a sua descendência, portanto permite ao melhorista inferir sobre o potencial dos indivíduos como genitores.

É interessante destacar o Modelo 1 que assume a estrutura de independência (ID) para modelar o efeito genético nos diferentes locais e colheitas. Esse resultado sugere que, mesmo quando se tem grandes desbalanceamento dos dados, em detrimento da abordagem de ANAVA, é possível realizar a análise conjunta de grupos de experimentos, sem a necessidade de retirar indivíduos que estão presentes apenas em alguns experimentos, mas não em todos. Em outras palavras, tais suposições são as mesmas empregadas pelo modelo fixo, muito embora o elevado grau de desbalanceamento não permita que tal análise seja utilizada.

Vale ressaltar que modelar estruturas de VCOV para efeitos relacionados nem sempre é considerado, mas trata-se de uma sutileza estatística para obter modelos mais realistas, com potencial preditivo para maximizar a acurácia das previsões. Nesse contexto, uso do modelos misto sem modelar estrutura de VCOV, em primeiro momento, pode e deve ser incorporado na rotina de análise dos programas de melhoramento genético de cana-de-açúcar. Certamente a análise conjunta grupos de experimentos pode ser realizada com vantagens, utilizando abordagem de modelos misto, e com o uso de modelo simples semelhante ao Modelo 1. O Modelo 1 é similar a abordagem tradicional de ANAVA que assume homogeneidade de variância para local e colheita, além de ausência de correlação genética entre os genótipos; portanto, não requer maiores conhecimentos estatísticos por parte dos usuários das análises.

A seleção do melhor modelo (dentro de algum critério) pode ser baseada em dois critérios de informação: o AIC e o BIC. Assim, em relação a decisão de seleção, para a maioria dos modelos, os dois critérios foram discordantes. Como exemplo tem-se o Modelo 19 ($FA1 \otimes UNST \otimes ID$,

com 102 parâmetros) que apresentou o menor valor de AIC (41426,59) e o Modelo 29 ($ID \otimes AR1 \otimes Kinship$, com 2 parâmetros) o menor valor de BIC (41885,83), ambos foram destacados com negrito (Tabela 6). Uma justificativa para esse resultado deve-se ao grande número de experimentos analisado. O conjunto de dados é composto de 44 experimentos, aqui também referidos como locais. A matriz de VCOV genética para local tem dimensão 44×44 , ou seja, é razoavelmente grande permitindo a presença de muitos parâmetros. Os modelos alternativos de matriz de VCOV para local possuem muitos parâmetros, consequentemente, foram mais penalizado pelo critério BIC. Verifica-se que em geral, o critério de informação BIC aplica uma maior penalidade para a matriz de local, priorizando o uso de estrutura mais simples, como a ID ou DIAG. Esse resultado evidencia que para esse grupo de experimentos, o uso desse critério pode não fazer sentido, uma vez que, a presença de heterogeneidade de variância e correlação genética entre locais é esperada. Esse argumento sugere que para este caso, particularmente, o AIC poderá ser o melhor critério de seleção, apesar de permitir mais parâmetros, aplicando menor penalidade, e portanto permitindo o uso de estruturas mais complexas, proporcionando o ajuste de modelos mais realistas para locais e colheitas.

Diante do exposto, para seleção do melhor modelo considerou-se apenas o critério AIC. Assim, de acordo com AIC verificou-se que o a estrutura FA1 foi a melhor para local. Outros estudos tem recomendado a estrutura Fator analítico para modelar a VCOV genética de local (THOMPSON et al., 2003; BOER et al., 2007; SMITH et al., 2007; MEYER, 2009). Já a matriz UNST foi a melhor para colheita. A matriz de VCOV genética para colheita tem dimensão 5×5 , requerendo a estimação de no máximo 15 parâmetros. É uma matriz pequena, e o ajuste do modelo UNST foi simples e rápido. Esse tipo de matriz captura toda a variação de forma geral, porém requer estimação do número máximo de parâmetros, o que a torna indesejável em alguns cenários. Assim, para colheita optou-se por outro modelo mais parcimonioso, definido como AR1 (autoregressiva heterogênea de primeira ordem). A estrutura AR1 captura a informação de correlação entre sucessivas colheitas, e explica de forma sistemática a dependência temporal existente. Nesse contexto, o conhecimento do padrão de resposta dos dados deve ser considerado durante a escolha do modelo adequado. A produtividade da cana-de-açúcar tende a reduzir ao longo das colheitas, e diante desse comportamento, espera-se, por exemplo, uma maior correlação entre colheitas mais próximas (por exemplo, corte 1 e 2), e menor entre as mais distante (por exemplo, corte 1 e 3), em função de alterações fisiológicos e genéticas. Nesse sentido, o Modelo 11 ($FA1 \otimes AR1 \otimes ID$, com 89 parâmetros), o segundo melhor de acordo com AIC, pode ser considerando o melhor modelo. É compatível os padrões de respostas observado no campo, na avaliação de grupos de experimentos de cana-de-açúcar. Como esperado, assume a matriz

FA1 para local e AR1 para colheita. Acredita-se que a estrutura AR1 não foi indicada pelo AIC como a melhor, devido esse critério não penalizar fortemente estruturas mais parametrizadas. Ao contrário do BIC, observa-se uma certa prioridade por estruturas mais parametrizadas, como esperado. Porém a matriz de VCOV para local foi penalizada fortemente por ambos critérios, o que resultou numa série de modelos que não rodou e foram identificados com “NR”.

Pelo exposto, assumiu-se que o Modelo 11, no geral, é o melhor modelo. É um modelo realista que assume estrutura adequadamente de VCOV para explicar as interações existentes entre genótipos, locais e corte. Este modelo poderá ser utilizado para predição do valor genotípico de genótipos de cana-de-açúcar em programas de melhoramentos. Os valores genotípicos quando ranqueados, poderão auxiliar o melhorista na identificação e seleção de genótipos superiores e de futuras variedades para recomendação comercial.

Vale salientar que um aspecto aparentemente incoerente do Modelo 11 é que, embora tenha sido considerado o melhor modelo, ele não inclui a informação de parentesco genético. E como já mencionado, de acordo com AIC, os modelos são “melhorados” quando a informação de parentesco é incorporada. Verifica-se então, que o Modelo 31 correspondente ao 11 com o parentesco (FA1 $\otimes$ AR1 $\otimes$ Kinship, com 89 parâmetros) não foi testado devido a limitações computacionais (Computador Dell Studio XPS 8100 Intel® Core™ i7-860, 17, 2.80GHz 16GB utilizando o Windows 7 Home Premium e o GenStat 14^o). Logo de início houve problemas com alocação de memória o que inviabilizou o início do processo iterativo. Mas, quando colocados lado a lado os modelos sem parentesco (1 - 20) e com parentesco (21 - 40), verifica-se que sempre o modelo correspondente com o parentesco foi melhor. Portanto, as evidências sugerem que se o Modelo 31 tivesse sido ajustado, provavelmente ele seria melhor modelo. O mesmo admite estruturas de VCOV genética (para local e colheitas) que explicam mais adequadamente a natureza da correlações existentes. E, adicionalmente, incorpora a informação de parentesco genético. Em resumo, trata-se de um modelo mais realista, e portanto, espera-se que possua maior poder preditivo e que seja capaz de gerar predições mais acuradas, o que resultaria numa maior eficiência de seleção. Espera-se que o ajuste de modelos semelhantes seja possível num futuro próximo, com o avanço dos algoritmos e capacidade computacional.

A Tabela 8 mostra o ranqueamento dos 74 genótipos em função das predições do valor genotípico (VG) de TPH para os Modelos 1 e 11, e a média aritmética para o Modelo Tradicional. Para fins de comparação o ordenamento dos genótipos deu-se em função do Modelo 11. Verifica-se que a classificação dos genótipos é discordante entre os modelos.

As comparações entre os modelos são mostradas também nas Figuras 8, 9 e 10. Determinou-se o coeficiente de correlação *Spearman* (ρ) para cada comparação, afim de medir a intensidade

Tabela 7 – Diferentes estruturas de variâncias e covariâncias $\mathbf{G}$ avaliadas, com os correspondentes números de parâmetros estimados n_{PAR} , e respectivos valores de AIC (Critério de Informação de Akaike) e BIC (Critério de Informação de Bayesiano)

Matrix $\mathbf{G}$	Modelo	n_{PAR} em $\mathbf{G}^\dagger$	AIC	BIC
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{I}_{I \times I}^G$	1) ID $\otimes$ ID $\otimes$ ID	$(1 + 1) - 1 = 1$	42473,64	42502,11
	2) DIAG $\otimes$ ID $\otimes$ ID	$(44 + 1) - 1 = 44$	42266,29	42600,77
	3) FA1 $\otimes$ ID $\otimes$ ID	$(88 + 1) - 1 = 88$	41660,41	42308,02
	4) UNST $\otimes$ ID $\otimes$ ID	$(990 + 1) - 1 = 990$	NR	NR
	5) ID $\otimes$ DIAG $\otimes$ ID	$(1 + 5) - 1 = 5$	42448,34	42505,27
	6) DIAG $\otimes$ DIAG $\otimes$ ID	$(44 + 5) - 1 = 48$	42257,09	42620,03
	7) FA1 $\otimes$ DIAG $\otimes$ ID	$(88 + 5) - 1 = 92$	41666,14	42342,22
	8) UNST $\otimes$ DIAG $\otimes$ ID	$(990 + 5) - 1 = 994$	NR	NR
	9) ID $\otimes$ AR1 $\otimes$ ID	$(1 + 2) - 1 = 2$	42298,04	42333,62
	10) DIAG $\otimes$ AR1 $\otimes$ ID	$(44 + 2) - 1 = 45$	41937,97	42279,56
	11) FA1 $\otimes$ AR1 $\otimes$ ID	$(88 + 2) - 1 = 89$	41487,76	42142,49
	12) UNST $\otimes$ AR1 $\otimes$ ID	$(990 + 2) - 1 = 991$	NR	NR
	13) ID $\otimes$ AR1 _{Het} $\otimes$ ID	$(1 + 6) - 1 = 6$	42269,51	42333,56
	14) DIAG $\otimes$ AR1 _{Het} $\otimes$ ID	$(44 + 6) - 1 = 49$	41922,11	42292,17
	15) FA1 $\otimes$ AR1 _{Het} $\otimes$ ID	$(88 + 6 + 1) - 1 = 94$	41489,18	42172,37
	16) UNST $\otimes$ AR1 _{Het} $\otimes$ ID	$(990 + 6) - 1 = 995$	NR	NR
	17) ID $\otimes$ UNST $\otimes$ ID	$(1 + 15) - 1 = 15$	42241,50	42369,60
	18) DIAG $\otimes$ UNST $\otimes$ ID	$(44 + 15) - 1 = 58$	41908,16	42342,27
	19) FA1 $\otimes$ UNST $\otimes$ ID	$(88 + 15) - 1 = 102$	41426,59	42173,83
	20) UNST $\otimes$ UNST $\otimes$ ID	$(990 + 15) - 1 = 1004$	NR	NR
$\mathbf{G} = \mathbf{G}_{J \times J}^L \otimes \mathbf{G}_{K \times K}^C \otimes \mathbf{A}_{I \times I}^G$	21) ID $\otimes$ ID $\otimes$ Kinship	$(1 + 1) - 1 = 1$	42177,28	42205,74
	22) DIAG $\otimes$ ID $\otimes$ Kinship	$(44 + 1) - 1 = 44$	42057,86	42392,34
	23) FA1 $\otimes$ ID $\otimes$ Kinship	$(88 + 1) - 1 = 88$	NR	NR
	24) UNST $\otimes$ ID $\otimes$ Kinship	$(990 + 1) - 1 = 990$	NR	NR
	25) ID $\otimes$ DIAG $\otimes$ Kinship	$(1 + 5) - 1 = 5$	42164,91	42221,85
	26) DIAG $\otimes$ DIAG $\otimes$ Kinship	$(44 + 5) - 1 = 48$	42049,15	42412,09
	27) FA1 $\otimes$ DIAG $\otimes$ Kinship	$(88 + 5) - 1 = 92$	NR	NR
	28) UNST $\otimes$ DIAG $\otimes$ Kinship	$(990 + 5) - 1 = 994$	NR	NR
	29) ID $\otimes$ AR1 $\otimes$ Kinship	$(1 + 2) - 1 = 2$	41850,25	41885,83
	30) DIAG $\otimes$ AR1 $\otimes$ Kinship	$(44 + 2) - 1 = 45$	41610,50	41952,09
	31) FA1 $\otimes$ AR1 $\otimes$ Kinship	$(88 + 2) - 1 = 89$	NR	NR
	32) UNST $\otimes$ AR1 $\otimes$ Kinship	$(990 + 2) - 1 = 991$	NR	NR
	33) ID $\otimes$ AR1 _{Het} $\otimes$ Kinship	$(1 + 6) - 1 = 6$	41834,69	41898,74
	34) DIAG $\otimes$ AR1 _{Het} $\otimes$ Kinship	$(44 + 6) - 1 = 49$	41592,66	41962,73
	35) FA1 $\otimes$ AR1 _{Het} $\otimes$ Kinship	$(88 + 6) - 1 = 93$	NR	NR
	36) UNST $\otimes$ AR1 _{Het} $\otimes$ Kinship	$(990 + 6) - 1 = 995$	NR	NR
	37) ID $\otimes$ UNST $\otimes$ Kinship	$(1 + 15) - 1 = 15$	41811,31	41939,41
	38) DIAG $\otimes$ UNST $\otimes$ Kinship	$(44 + 15) - 1 = 58$	41582,15	42016,26
	39) FA1 $\otimes$ UNST $\otimes$ Kinship	$(88 + 15) - 1 = 102$	NR	NR
	40) UNST $\otimes$ UNST $\otimes$ Kinship	$(990 + 15) - 1 = 1004$	NR	NR

$\mathbf{G}$: é a matriz de VCOV genética; ID: Independente; DIAG: Diagonal; FA1: Fator analítico de primeira ordem; AR1: Auto regressiva de primeira ordem; AR1_{Het}: Auto regressiva de primeira ordem (heterogêneo) e UNST: não-estruturado. Os menores valores dos critérios AIC e BIC estão destacados em negrito, indicam o melhor modelo. †: número total da matriz de VCOV do modelo.

da associação entre os modelos em função das coincidências de ranqueamento. A Figura 8 mostra a comparação entre o Modelo 1 e o Tradicional. Os valores genotípico preditos obtidos com o Modelo 1 e as médias estimadas de TPH com o Modelo Tradicional foram correlacionados significativamente ($\rho = 0,72$, p-valor < 0,001).

A Figura 9 mostra a comparação entre o Modelo 11 e o Tradicional. Nesta os valores genotípicos preditos com o Modelo 11 foram correlacionado significativamente com as médias estimadas para TPH com o modelo Tradicional ($\rho = 0,63$, p-valor < 0,001). Há coincidência de ordenamento dos genótipos, porém o melhoramento de cana requer maior eficiência para se-

leção dos melhores indivíduos, nesse contexto essa associação não é classificada como alta. A associação existente não é suficiente para declarar o Modelo Tradicional uma boa opção para o melhoramento. Há ganho de eficiência quando se utiliza o Modelo 11, e esse ganho não é refletido com o Modelo Tradicional, portanto, o Modelo Tradicional não substitui com eficiência do 11.

Já a Figura 10 tem-se a comparação do Modelo 1 e 11. A correlação entre os valores genotípicos preditos dos dois modelos foi alta e significativa ($\rho = 0,87$, $p\text{-valor} < 0,001$). Isso sugere que existe concordância de ordenamento entre os dois modelos. E quando se observa o ponto de corte, existem cinco genótipos (50%) que foram selecionados com o Modelo 11, mas não foram com o Modelo 1. O argumento a esse resultado é que quando a matriz de VCOV genética é modelada adequadamente, as correlações genéticas entre locais e entre colheitas são consideradas, o modelo torna-se mais realista e aumenta sua capacidade preditiva.

Ainda nas Figuras 8, 9 e 10, notam-se que as médias estimadas observadas tiveram maior amplitude de variação do que os valores preditos. Isso é justificado pela propriedade de encolhimento do preditor, típica da análise de modelos mistos, que é desejável pois penaliza observações discrepantes (PIEPHO et al., 2008).

Tabela 8 – Ranqueamento dos 74 genótipos em função das predições do valor genotípico médio de TPH (VG) obtidas pelos modelos 11 e 1, e da média aritmética de TPH estimada pelo modelo tradicional (Continua)

Genótipos	Modelo 11		Modelo 1		Modelo Tradicional	
	VG	Ranking	VG	Ranking	Média de TPH	Ranking
SP86155	15,700	1 °	12,900	5 °	16,745	8 °
RB867515	15,580	2 °	12,750	14 °	14,785	19 °
IAC911099	14,830	3 °	13,330	2 °	16,372	10 °
RB925345	14,790	4 °	12,720	16 °	18,607	1 °
IACSP933046	14,770	5 °	12,860	9 °	15,410	15 °
IACSP944004	14,730	6 °	12,910	4 °	15,900	12 °
SP841431	14,670	7 °	12,710	17 °	14,503	21 °
SP891115	14,480	8 °	12,890	6 °	14,465	23 °
RB825336	14,410	9 °	12,750	14 °	17,422	4 °
SP832847	14,310	10 °	12,670	20 °	11,770	61 °
IACSP942094	14,270	11 °	12,760	13 °	14,273	26 °
IACSP933050	14,200	12 °	12,860	9 °	15,356	17 °
RB855536	14,110	13 °	12,700	18 °	17,794	3 °
IACSP955011	14,000	14 °	12,820	11 °	13,925	34 °
IACSP955000	13,970	15 °	12,880	7 °	14,270	27 °
IACSP946033	13,950	16 °	12,670	20 °	14,446	24 °
SP911049	13,770	17 °	12,700	18 °	12,741	51 °
RB72454	13,690	18 °	13,370	1 °	15,409	16 °
IAC873396	13,630	19 °	12,960	3 °	16,667	9 °
IACSP966021	13,540	20 °	12,860	9 °	14,192	29 °
IACSP953018	13,480	21 °	12,680	19 °	13,057	47 °
SP813250	13,450	22 °	12,870	8 °	13,934	32 °
RB835054	13,430	23 °	12,670	20 °	14,467	22 °
IACSP943591	13,330	24 °	12,750	14 °	13,688	39 °
IACSP953264	13,330	24 °	12,610	26 °	12,666	53 °
IAC823092	13,320	25 °	12,810	12 °	16,813	7 °
IACSP932060	13,290	26 °	12,820	11 °	14,003	30 °
RB855453	13,280	27 °	12,660	21 °	13,229	43 °
IAC822045	13,260	28 °	12,830	10 °	16,964	6 °
IACSP946025	13,200	29 °	12,730	15 °	13,892	35 °
SP801842	13,200	30 °	12,900	5 °	15,688	13 °
IACSP942101	13,190	31 °	12,710	17 °	13,714	38 °
IACSP945041	13,130	32 °	12,650	22 °	12,977	49 °
IAC873184	13,120	33 °	12,830	10 °	16,978	5 °
SP791011	13,110	34 °	12,650	22 °	13,327	41 °
IACSP931020	13,100	35 °	12,620	25 °	13,117	46 °
RB835089	13,070	36 °	12,650	22 °	18,272	2 °
RB855156	12,980	37 °	12,600	27 °	12,844	50 °
IACSP952213	12,960	38 °	12,590	28 °	12,509	56 °
SP775181	12,950	39 °	12,630	24 °	13,833	36 °
IAC873187	12,800	40 °	12,720	16 °	16,101	11 °
SP835073	12,800	40 °	12,590	28 °	14,762	20 °
IACSP956087	12,780	41 °	12,730	15 °	13,959	31 °
IACSP953104	12,770	42 °	12,500	34 °	12,530	55 °
IACSP953028	12,760	43 °	12,630	24 °	13,146	45 °
IACSP955050	12,710	44 °	12,550	31 °	12,261	58 °
PAV9409	12,660	45 °	12,640	23 °	7,689	74 °
IACSP955048	12,650	46 °	12,510	33 °	12,030	59 °
IACSP955110	12,650	46 °	12,490	35 °	11,926	60 °
IAC832285	12,640	47 °	12,650	22 °	15,586	14 °
IACSP943581	12,600	48 °	12,630	24 °	13,158	44 °
RB835486	12,490	49 °	12,170	49 °	12,983	48 °
IACSP942111	12,470	50 °	12,480	36 °	12,423	57 °
SP801816	12,450	51 °	12,620	25 °	9,687	71 °
IACSP944002	12,400	52 °	12,620	25 °	12,704	52 °
IAC811032	12,260	53 °	12,570	30 °	14,937	18 °
IACSP956114	12,060	54 °	12,510	33 °	11,469	66 °
IACSP942180	12,040	55 °	12,330	43 °	11,726	62 °
IACSP952288	12,030	56 °	12,400	41 °	11,395	67 °
PO861107	11,880	57 °	12,470	37 °	14,193	28 °
IAC862210	11,690	58 °	12,580	29 °	14,444	25 °
IACSP956100	11,640	59 °	12,520	32 °	11,620	64 °
IAC911206	11,500	60 °	12,210	48 °	11,212	68 °

Tabela 8 – Ranqueamento dos 74 genótipos em função das previsões do valor genotípico médio de TPH (VG) obtidas pelos modelos 11 e 1, e da média aritmética de TPH estimada pelo modelo tradicional (Conclusão)

Genótipos	Modelo 11		Modelo 1		Modelo Tradicional	
	VG	Ranking	VG	Ranking	Média de TPH	Ranking
IAC911121	11,490	61 ^o	12,290	45 ^o	11,568	65 ^o
IACSP966026	11,100	62 ^o	12,320	44 ^o	11,703	63 ^o
IAC832396	11,100	62 ^o	12,400	41 ^o	13,631	40 ^o
IACSP945003	11,050	63 ^o	12,410	40 ^o	10,294	69 ^o
IAC834107	10,960	64 ^o	12,420	39 ^o	13,794	37 ^o
IAC831313	10,930	65 ^o	12,440	38 ^o	13,926	33 ^o
IACSP946010	10,870	66 ^o	12,400	41 ^o	10,253	70 ^o
IAC832405	10,790	67 ^o	12,360	42 ^o	13,278	42 ^o
IAC853229	10,530	68 ^o	12,270	46 ^o	12,604	54 ^o
IACSP945072	10,440	69 ^o	12,320	44 ^o	9,386	72 ^o
IACSP936035	10,260	70 ^o	12,250	47 ^o	8,568	73 ^o

Figura 8 – Gráfico das previsões do valor genético de TPH do Modelo 1 contra as médias estimadas de TPH do Modelo Tradicional para 74 genótipos ($\rho = 0,72$, $p\text{-valor} < 0,001$), com informações dos 44 experimentos

Figura 9 – Gráfico das previsões do valor genético de TPH do Modelo 11 contra as médias estimadas de TPH do Modelo Tradicional para 74 genótipos ($\rho = 0,63$, $p\text{-valor} < 0,001$), com informações dos 44 experimentos

Figura 10 – Gráfico das previsões do valor genético de TPH do Modelo 1 contra Modelo 11 para 74 genótipos ($\rho = 0,87$, $p\text{-valor} < 0,001$), com informações dos 44 experimentos

5 DISCUSSÃO

A decisão de identificar genótipos superiores em grupos de experimentos de cana-de-açúcar muitas vezes está limitada a uma medida estatística não robusta, a média aritmética e/ou a análise conjunta de variância com descarte de indivíduos. A média aritmética, apesar de ignorar o delineamento, tem sido a medida estatística utilizada em alguns casos para estimar o potencial produtivo dos genótipos. Trata-se de uma medida resumo extremamente sensível a presença de observações discrepantes. Como foi observado, valores extremos de TPH podem inflacionar a média, subestimando ou superestimando o perfil produtivo dos genótipos. Outra questão é que diante do desbalanceamento existente, é comum genótipos terem números diferentes de observações, visto que, não necessariamente todos foram avaliados em todos locais e/ou tiveram mesmo número de colheitas. Isto é típico da rotina dos programas de melhoramento. Nesse caso, espera-se que a produtividade média de um genótipo seja mais consistente quando estimada com maior número de observações. Poucas observações não reflete o real potencial produtivo do indivíduo. Este fato, aliado a presença de valores discrepantes, pode levar a obtenção de estimativas imprecisas. Nessas condições, o uso da média com objetivo de comparação e ranqueamento dos genótipos não é adequada e deve ser evitada. A mediana é outro recurso estatístico que dá uma ideia da distribuição dos dados, com a vantagem de não ser afetada por valores extremos. Porém, é menos informativa que a média, não sendo portanto uma boa alternativa. Contudo, pode ser enganoso assumir que a estimativa da produtividade via média aritmética seja um valor representativo de todas observações. Essa informação não é suficiente por si só para caracterizar a produtividade de um genótipo.

A aplicação de modelos misto é uma alternativa aos modelos tradicionais utilizado nos programas de melhoramento. Numa situação de grande desbalanceamento, os dados podem ser analisados facilmente, e grupos de efeitos podem ser assumidos como aleatórios ao invés de fixos (SMITH et al., 2001; PIEPHO et al., 2008). Um argumento contra o uso de modelos misto para o melhoramento de plantas era a falta de software adequado. Mas, atualmente, maiores recursos computacionais estão disponíveis possibilitando o uso de ferramentas estatísticas mais eficientes, bem como softwares apropriados.

Verificou-se que o modelo misto oferece uma adequada abordagem para análise conjunta de grupos de experimentos de cana-de-açúcar. É uma solução para análise de dados desbalanceados. Substitui com vantagens o modelo tradicional de médias e a abordagem clássica de análise de variância. Quando se tem desbalanceamento até mesmo aplicação de modelo misto mais

simples (assumindo a identidade para a matriz de VCOV de locais e colheitas), possui muitas vantagens. É possível realizar a análise conjunta dos experimentos sem descartar indivíduos e estabelecer um ranqueamento em função de valores genéticos preditos. Para comparação de genótipos, tais valores são estimativas mais precisas e acuradas do que a média aritmética.

Demonstrou-se aqui que é possível ajustar modelo mais realistas, que admitam a presença de correlação entre as observações, o que permite modelar apropriadamente uma matriz de VCOV para o efeito de interação $G \times L \times C$. A exploração da informação de interação é essencial para assegurar o correto ordenamento dos indivíduos. O modelo misto com seleção de matriz de VCOV ajusta os dados de TPH melhor que o modelo tradicional. Em princípio, a escolha dos modelos de VCOV para o efeito de locais, requer o conhecimento prévio da distribuição espacial dos experimentos. Para colheita, é importante haver uma compreensão da resposta biológica da espécie sendo estudada. Diante das considerações, tem-se que a distribuição geográfica dos experimentos do IAC está concentrada no estado de São Paulo, especialmente na região norte, com alguns ensaios no Triângulo Mineiro e em Goiás. As diferentes condições edafoclimáticas dos locais alertam para a possibilidade de ocorrência de alterações de ordenamento dos genótipos. É previsível a existência de interação complexa, e esta pode acarretar em mudanças no ranqueamento. Nesse contexto, a matriz fator analítico (FA) mostrou-se adequada para modelar a estrutura de VCOV para efeito de interação $G \times L$.

Entre os 40 modelos testados, o modelo proposto assumiu a matriz fator analítico de primeira ordem (FA1) para efeito genético em locais e autoregressiva de primeira ordem (AR1) para efeito genético em colheitas. Ambas foram selecionadas conjuntamente como as melhores matrizes baseado-se no critério AIC e na coerência do padrão de respostas dos dados. Tais estruturas por se ajustarem melhor ao padrão natural de resposta das observações, substitui com vantagens a matriz de independência assumida pelo modelo de ANAVA.

A matriz FA atribuída aos efeitos genéticos em local tem sido sugerida na literatura (KELLY et al., 2007; BURGUEÑO et al., 2011; SO e EDWARDS, 2011). Sua configuração descreve adequadamente a natureza da interação $G \times L$, modelando a heterogeneidade de variância entre 44 experimentos e fornecendo estimativas de associações genéticas entre os pares de locais. Tal estrutura captura a variação de comportamento dos genótipos no local e entre locais. Em análise de grupos de experimentos de cana-de-açúcar, quando tal estrutura é empregada, a mesma contribui para obtenção de predições mais acuradas, aumentando assim, a eficiência de seleção. Isso também foi evidenciado por Kelly et al. (2007). Na Austrália, os programas de melhoramento de plantas tem se beneficiado com a aplicação desse modelo (KELLY et al., 2007). Fundamentalmente é assegurado que a estrutura genética seja modelada com modelos mais realistas.

Similarmente, a matriz AR ajustada para efeito genético de colheitas, confirma existência de dependência temporal entre as colheitas sucessivas (planta, soca e ressoca). Essa dependência é modelada de forma que a correlação é maior entre colheitas mais próximas, corte 1 e 2 ou corte 2 e 3, por exemplo, e tende a diminuir entre colheitas mais distante. Trata-se portanto, de um clássico exemplo de dados longitudinais. Modelar esse padrão de resposta é essencial, uma vez que, provavelmente, fatores fisiológico e bioquímicos estejam diretamente relacionado a esse comportamento. E ainda, pode-se esperar que a expressão de determinado gene(s) possa estar associada a diferentes condições fisiológicas da cana.

Em resumo, o modelo final resultante foi o Modelo 11 ($FA1 \otimes AR1 \otimes ID$). Produz previsões do valor genotípico médio de TPH (VG) para cada indivíduo, que são medidas mais acurada da produtividade média. A aplicação desse modelo, poderá proporcionar ganhos de eficiência, na identificação e seleção de genótipos superiores. Os resultados verificados aqui permitem sugerir seu uso nos programas de melhoramento de cana-de-açúcar.

Ao assumir que a média aritmética possa ser empregada, está sendo feita um enorme simplificação do que ocorre nas rotinas dos programas de melhoramento. Porém, o cenário aqui avaliado considerou uma situação de análises não supervisionadas, ou seja, sem avaliação dos resultados por um especialista. Nesse caso, a abordagem usual seria de fato a análise de variância com modelo fixo, com cálculo das médias aritméticas. O emprego de modelos mistos, combinado com a expertise dos melhoristas, compõe o cenário ideal, e em realidade é isso que está sendo sugerido aqui. Como os dados já são usualmente obtidos, o custo em adquirir softwares é mínimo perto do investimento já realizado, e o benefício seria grande, já que os dados dos experimentos têm valiosas informações.

O Modelo 11 não considera as relações existente entre os indivíduos. De forma geral, os modelos com inclusão de parentesco, foram melhores quando comparados com seus correspondentes sem esta informação, de acordo o critério AIC. Isso reforça que modelos com parentesco são mais realistas, e portanto produzem estimativas mais acuradas do valor genético (VA), e por seguinte, do valor genotípico (VG). Piepho et al. (2008) demonstrou que o uso de modelos mistos incluindo a informação de parentesco produz estimativas mais acuradas quando comparada aos modelos sem essa informação. Isso também foi demonstrado por Wey e Borralho (2000), Purba et al. (2001) e Furlani et al. (2005). Essa superioridade é justificada pela capacidade dos preditores usar a informação dos indivíduos através da correlação genética que surge da genealogia. Os modelos com a informação de parentesco produzem previsões que computam a informação de variância genética aditiva, e em cana-de-açúcar, essa informação, semelhante ao melhoramento animal, pode ser aproveitada para definir genitores promissores para futuros

cruzamentos.

Aqui, os modelos com informação de parentesco, são uma extensão dos modelos sem essa informação. As estruturas de VCOV para locais e colheitas foram mantidas, e a matriz de independência entre os genótipos foi substituída pela matriz de parentesco genético aditivo, **A**. Porém, algumas pesquisas tem evidenciado que a não inclusão de genealogias completas resulta em modelos que geram estimativas viesada do valor genético (PIEPHO et al., 2008; MEHRABANI-YEGANEH et al., 1999; DUREL et al., 1998). Ademais, pesquisadores da Austrália (ATKIN; DIETERS; STRINGER, 2009) estudaram a extensão do impacto da quantidade de informação de genealogia de cana-de-açúcar, sobre a acurácia da estimativas de componentes de variância e do valor genético. Os autores mostraram que quando a informação de parentesco é mínima, são fornecidas estimativas viesadas de variância aditiva e consequentemente do valor genético. A acurácia da seleção de genitores torna-se baixa quando comparada aos modelos com genealogia completa.

Porém, o cenário atual permite facilmente resolver este problema. Há atualmente diversos tipos de marcadores moleculares que podem ser empregados com custo bastante razoável. Lima et al. (2002) mostraram que existe correlação entre a similaridade genética determinada por meio de marcadores AFLP e por genealogia. Portanto o emprego de marcadores moleculares pode ajudar a determinar com mais acurácia o grau de parentesco entre os genótipos.

Embora os dados sugiram que a incorporação do parentesco torne o modelo mais realista, limitações computacionais ainda existem, o que inviabilizou o ajuste de tal modelo. Maiores recursos ainda são necessários. Apesar desse impasse, o Modelo 11 possui atributos que garante uma boa capacidade preditiva. Seu emprego certamente poderá contribuir para o aumento de eficiência de seleção de genótipos superiores em programas de melhoramento de cana-de-açúcar, aliada evidentemente à experiência dos melhoristas.

6 CONCLUSÃO

O modelo misto aqui apresentado pode contribuir para a identificação de genótipos superiores de cana-de-açúcar. Aliado à indispensável experiência dos melhoristas, pode contribuir com os programas de melhoramento genético.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **The Institute of Electrical and Electronics Engineers Transaction and Automatic Control**, Notre Dame, v. 19, p. 716-723, 1974.

ALLARD, R.W. **Principles of plant breeding**. 2nd. ed. New York: John Wiley & Sons, 1999. 254 p.

ATKIN, F.C.; DIETERS, M.J.; STRINGER, J.K. Impact of depth of pedigree and inclusion of historical data on the estimation of additive variance and breeding values in a sugarcane breeding program. **Theoretical and Applied Genetics**, New York, v. 119, p. 555-565, 2009.

BALZARINI, M. Applications of mixed models in plant breeding. In: KANG, M.S. **Quantitative genetics, genomics and plant breeding**. New York: CABI Publishing, 2001. p. 353-363.

BARBOSA, M.H.P.; RESENDE, M.D.V.; PETERNELLI, L.A.; BRESSIANI, J.A.; SILVEIRA, L.C.I.; SILVA, F.L.; FIGUEIREDO, I.C.R. Use of REML-BLUP for the selection of sugarcane families specialized in biomass production. **Crop Breeding and Applied Biotechnology**, Londrina, v. 4, p. 218-226, 2004.

BARBOSA, M.H.P.; RESENDE, M.D.V.; DIAS, L.A.S.; BARBOSA, G.V.S.; OLIVEIRA, R. A.; PETERNELLI, L. A.; DAROS, E. Genetic improvement of sugarcane for bioenergy: the Brazilian experience in network research with RIDESA. **Crop Breeding and Applied Biotechnology**, Londrina, v. 2, p. 87-98, 2012.

BAUER, A.M.; LÉON, J. Multiple-trait breeding values for parental selection in self-pollinating crops. **Theoretical and Applied Genetics**, New York, v. 116, p. 235-242, 2008.

BERNARDO, R. Prediction of maize single-cross performance using RFLPs and information from related hybrids. **Crop Science**, Madison, v. 34, p. 20-25, 1994.

BERNARDO, R.; Yu J.M. Prospects for genomewide selection for quantitative traits in maize. **Crop Science**, Madison, v. 47, p. 1082-1090, 2007.

BERNARDO, R. **Breeding for quantitative traits in plants**. 2nd. ed. Minnesota: Stemma Press, 2010, 400p.

BLANCO, I.A.D.; GLAZ, B.; EDMÉ, S.J. Improving efficiency of sugarcane genotype selection in Florida. **Crop science**, Madison, v. 50, p. 1744-1753, 2010.

- BOER, M. P.; WRIGHT, D.; FENG, L.; PODLICH, D. W.; LUO, L.; COOPER, M.; EEUWIJK, F.A. van A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. **Genetics**, Bethesda, v. 177, p. 1801-1813, 2007.
- BOHN, M.; GROH, S.; KHAIRALLAH, M.M.; HOISINGTON, D.A.; UTZ, H.F.; MELCHINGER, A.E. Re-evaluation of the prospects of marker-assisted selection for improving insect resistance against *Diatraea* spp. in tropical maize by cross validation and independent validation. **Theoretical and Applied Genetics**, New York, v. 103, p. 1059-1067, 2001.
- BONNETT, D.G.; REBETZKE, G.J.; SPIELMEYER, W. Strategies for efficient implementation of molecular markers in wheat breeding. **Molecular Breeding**, Berlin, v. 15, p. 75-85, 2005.
- BOZDOGAN, H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. **Psychometrika**, Colorado, v. 52, p. 345-370, 1987.
- BRESEGHELLO, F.; SORRELLS, M.E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. **Genetics**, Bethesda, v. 172, p. 1165-1177, 2006.
- BROWN, H.; PRESCOTT, R. **Applied Mixed Models in Medicine**. New York: John Wiley & Sons, 2006. 478 p.
- BROWN, A.H.D.; DANIELS, J.; LATTE, B.D.H. Quantitative genetics of sugarcane. **Theoretical and Applied Genetics**, New York, v. 38, p. 361-369, 1968.
- BRUMMER E.C. Applying genomics to alfalfa breeding programs. **Crop Science**, Madison, v. 44, p. 1904-1907, 2004.
- BUCKERIDGE, M.S.; SOUZA, A.P.; ARUNDALE, R.A.; ANDERSON-TEIXEIRA, K.J.; DELUCIA, E. Ethanol from sugarcane in Brazil: a 'midway' strategy for increasing ethanol production while maximizing environmental benefits. **GCB Bioenergy**, Urbana, v. 4, p. 119-126, 2012.
- BURGUEÑO, J.; CROSSA, J.; CORNELIUS, P.L.; TRETHOWAN, R.; McLAREN, G.; KRISHNAMACHARI, A. Modeling additive $\times$ environment and additive $\times$ additive $\times$ environment using genetic covariances of relatives of wheat genotypes. **Crop Science**, Madison, v. 43, p. 311-320, 2007.
- BURGUEÑO, J.; CROSSA, J.; CORNELIUS, P.L.; YANG, R.C. Using factor analytic models for joining environments and genotypes without crossover genotype $\times$ environment interaction. **Crop Science**, Madison, v. 48, p. 1291-1305, 2008.
- BURGUEÑO, J.; CROSSA, J.; COTES, J.M.; VICENTE, F.S.; DAS, B. Prediction assessment of linear mixed models for multi-environment trials. **Crop Science**, Madison, v. 51, p. 944-954, 2011.

BURGUEÑO, J.; DE LOS CAMPOS, G.; WEIGEL, K.; CROSSA, J. Genomic prediction of breeding values when modeling genotype× environment interaction using pedigree and dense molecular markers. **Crop Science**, Madison, v. 52, p. 707-719, 2012.

BURNHAM, K.P.; ANDERSON, D.R. Multimodel inference understanding AIC and BIC in model selection. **Sociological methods & research**, London, v. 33, p. 261-304, 2004.

CARLSON, J.P.; CHRISTIAN, L.L.; ROTHSCILD, M.F. et al. An evaluation of four procedures to rank centrally tested boars. **Journal of Animal Science**, Albany, v. 59, p. 934-940, 1984.

CHRISTENSEN, R.; PEARSON, L. M. Johnson, W. Case-Deletion Diagnostics for Mixed Models. **Technometrics**, Alexandria, v. 34, p. 38-45, 1992.

COLLARD, B.C.; MACKILL, D.J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. **Philosophical Transactions of the Royal Society B: Biological Sciences**, London, v. 363, p. 557-572, 2008.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira - Cana-de-açúcar: Safra 2012/2013, Segundo levantamento** - Agosto de 2012. Disponível em: <<http://www.conab.gov.br>>. Acesso em: 09 out. 2012.

CONSELHO DOS PRODUTORES DE CANA-DE-AÇÚCAR, AÇÚCAR E ÁLCOOL DO ESTADO DE SÃO PAULO. São Paulo, 2006. 54 p.

COOPER, M.; DELACY, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. **Theoretical and Applied Genetics**, New York, v. 88, p. 561-572, 1994.

CARDONA, C.A.; QUINTERO, C.A.; PAZ, I.C. Production of bioethanol from sugarcane bagasse: Status and perspectives. **Bioresource Technology**, Essex, v. 101, p. 4754-4766, 2010.

COWEN, N.M.; FREY, K.J. Relationships between three measures of genetic distance and breeding behaviour in oats (*Avena sativa* L.). **Genome**, Ottawa, v. 29, p. 97-106, 1987.

CROSSA, J. Statistical analyses of multilocation trials. **Advances in agronomy**, Mexico, v. 44, p. 55-85, 1990.

CROSSA, J.; FRANCO, J. Statistical methods for classifying genotypes. **Euphytica**, Wageningen, v. 137, p. 19-37, 2004.

CROSSA, J.; YANG, R.C.; CORNELIUS, P.L. Studying crossover genotype × environment interaction using linear-bilinear models and mixed models. **Journal of agricultural, biological, and environmental statistics**, Alexandria, v. 9, p. 362-380, 2004.

CROSSA, J.; BURGUEÑO, J.; CORNELIUS, P.L.; MCLAREN, G.; TRETHOWAN, R.; KRISHNAMACHARI, A. Modeling genotype $\times$ environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. **Crop Science**, Madison, v. 46, p. 1722-1733, 2006.

CROSSA, J.; BURGUEÑO, J.; DREISIGACKER, S.; VARGAS, M.; HERRERA-FOESSEL, S.A.; LILLEMO, M.; SINGH, R.P.; TRETHOWAN, R.; WARBURTON, M.; FRANCO, J.; REYNOLDS, M.; CROUCH, J. H.; ORTIZ, R. Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. **Genetics**, Bethesda, v. 177, p. 1889-1913, 2007.

CROSSA, J.; DE LOS CAMPOS, G.; PÉREZ, P.; GIANOLA, D.; BURGUEÑO, J.; ARAUS, J.L.; MAKUMBI, D.; SINGH, R.P.; DREISIGACKER, S.; YAN, J.; ARIEF, V.; BANZIGER, M.; BRAUN, H.J. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, Bethesda, v.186, p. 713-724, 2010.

DEKKERS J.C.M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, New York, v. 3, p. 22-32, 2002.

DEMPFLE, L. Problems in the Use of the Relationship Matrix in Animal Breeding. In: GIANOLA, D; HAMMOND, K. **Advances in statistical methods for genetic improvement of livestock**. New York: Springer, 1990. p. 454-474.

DEMPSTER A.; LAIRD, N.; RUBIN, D. Maximum likelihood from incomplete data via the EM Algorithm. **Journal of the Royal Statistical Society. Series B**, Oxford, v. 39, p. 1-38, 1977.

DIGGLE, P.J. An Approach to the Analysis of Repeated Measurements. **Biometrics**, Washington, v. 44, p. 959-971, 1988.

DUARTE, J.B; VENCOSKY, R. **Interação genótipos $\times$ ambientes: Uma Introdução à Análise “AMMI”**. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60p. (Série Monografias).

EBERHART, S.T.; RUSSELL, W.A. Stability parameters for comparing varieties. **Crop Science**, Madison, v. 6, p. 36-40, 1966.

ECKERMAN, P.J.; VERBYLA, A.P.; CULLIS, B.R.; THOMPSON, R. The analysis of quantitative traits in wheat mapping populations. **Australian Journal of Agricultural Research**, Melbourne, v. 52, p. 1195-1206, 2001.

FALCONER, D.S. The problem of environment and selection. **American Naturalist**, Chicago, v. 86, p. 293-298, 1952.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th. Ed. Essex, UK: Longman, 1996. 464 p.

FINLAY, K.W.; WILKINSON, G.N. The analysis of adaptation in a plant-breeding programme. **Australian Journal of Agricultural Research**, Collingwood, v. 14, p. 742-754, 1963.

FISHER, R.A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. **Transactions of the Royal Society of Edinburgh**, Edinburgh, v. 52, p. 399-433, 1918.

FISHER, R.A.; MACKENZIE, W.A. Studies in crop variation. II. The manurial response of different potato varieties. **The Journal of Agricultural Science**, Cambridge, v. 13, p. 311-320, 1923.

FLINT-GARCIA, S.A.; DARRAH, L.L.; McMULLEN, M.D.; HIBBARD, B.E. Phenotypic versus marker-assisted selection for stalk strength and second-generation European corn borer resistance in maize. **Theoretical and Applied Genetics**, New York, v. 107, p. 1331-1336, 2003.

FURLANI, R.C.M.; MORAES M.L.T.; RESENDE, M.D.V.; FURLANI, E. Jr.; GONÇALVES, P.; FILHO, W.V.V.; PAIVA, J.R. Estimation of variance components and prediction of breeding values in rubber tree breeding using the REML/BLUP procedure. **Genetics and Molecular Biology**, Ribeirão Preto, v. 28, p. 271-276, 2005.

GALWEY, N.W. **Introduction to mixed modelling**: beyond regression and analysis of variance. Chichester: Wiley, 2007. 376 p.

GAUCH, H.G. Model selection and validation for yield trials with interaction. **Biometrics**, Washington, v. 44, p. 705-715, 1988.

GAUCH, H.G. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. Amsterdam: Elsevier, 1992. 172 p.

GILMOUR, A.R.; GOGEL, B.J.; CULLIS, B.R.; THOMPSON, R. ASReml User Guide Release 3.0, 2009. Hernal Hempstead: VSN International. Disponível em: <<http://www.vsnl.co.uk>>. Acesso em: 01 Jun. 2011.

GOLLOB, H.F. A statistical model which combines features of factor analytic and analysis of variance techniques. **Psychometrika**, Baltimore, v. 33, p. 73-115, 1968.

GUERIN, L.; STROUP, W.W. A Simulation Study to Evaluate PROC MIXED Analysis of Repeated Measures Data. In: PROCEEDINGS OF THE 12TH ANNUAL CONFERENCE ON APPLIED STATISTICS IN AGRICULTURE, 2000, Manhattan. **Proceedings...** Manhattan: Kansas State University, 2000. p.170-203.

HENDERSON C.R.; KEMPTHORNE, O.; SEARLE, S.R.; Von KROSIGK, C.M. The estimation of environmental and genetic trends from records subject to culling. **Biometrics**, Washington, v. 13, p. 192-218, 1959.

HENDERSON, C.R. Best Linear Unbiased Estimation and Prediction under a Selection Model. **Biometrics**, Washington, v. 31, p. 423-447, 1975.

HENDERSON, C.R. A simple method for computing the inverse of a numerator relationship Matrix Used in prediction of breeding values. **Biometrics**, Washington, v. 32, p. 69-83, 1976.

HENDERSON, C.R. **Applications of linear models in animal breeding**. Ontario: University of Guelph, 1984. 462 p.

HENDERSON, C.R. Estimation of Variances in Animal Model and Reduced Animal Model for Single Traits and Single Records. **Journal of Animal Science**, Albany, v. 69, p. 1394-1402, 1986.

HOGARTH, D.M.; WU, K.K.; HEINZ, D.J. Estimating genetic variance in sugarcane using a factorial cross design. **Crop Science**, Madison, v. 21, p. 21-25, 1981.

HOSPITAL, F. Challenges for effective marker-assisted selection in plants. **Genetica**, Gravenhage, v. 136, p. 303-310, 2009.

HUANG, X.; WEI, X.; SANG, T.; ZHAO, Q.; FENG, Q.; ZHAO, Y. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. **Nature Genetics**, New York, v. 42, p. 961-967, 2010.

HUNTER, R.B. ASTA approach on minimum distance. In: WILKINSON, D. (Ed.) **44th Annual Corn Sorghum Industrial Research Conference**. Washington: Am Seed Trade Assoc, 1989. p. 193-195.

HU, X.; SPILKE, J. Variance-covariance structure and its influence on variety assessment in regional crop trials. **Field Crops Research**, Amsterdam, v. 120, p. 1-8, 2011.

JACKSON, P.A. Breeding for improved sugar content in sugarcane. **Field Crops Research**, Amsterdam, v. 92, p. 277-290, 2005.

JANNINK, J.L.; LORENZ, A.J.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, Oxford, v. 9, p. 166-177, 2010.

KANG M.S.; MILLER, J.D.; TAI PY P. Genetic and phenotypic path analyses and heritability in sugarcane. **Crop Science**, Madison, v. 23, p. 643-647, 1983.

KELLY, A.M.; SMITH, A.B.; ECCLESTON, J.A.; CULLIS, B.R. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. **Crop Science**, Madison, v. 47, p. 1063-1070, 2007.

KEMPTHORNE, O. **An introduction to genetic statistics**. Iowa: The Iowa State University, 1969. 545 p.

KEMPTON, R.A. The use of biplots in interpreting variety by environment interactions. **The Journal of Agricultural Science**, Cambridge, v. 103, p. 123-135, 1984.

KENNEDY, B.W.; MOXLEY, J.E. Comparison of genetic group and relationship methods for mixed model sire evaluation. **Journal of Dairy Science**, Lancaster, v. 58, p. 1507-1514, 1975.

KERR, R.J.; LI, L.; TIER, B.; DUTKOWSKI, G.W.; McRAE, T.A. Use of the numerator relationship matrix in genetic analysis of autopolyploid species. **Theoretical and Applied Genetics**, New York, v. 124, p. 1271-1282, 2012.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, Bethesda, v. 124, p. 743-756, 1990.

LANGE, K. **Mathematical and Statistical Methods for Genetic Analysis**. New York: Springer-Verlag, 1997. 367 p.

LATOUR, D.; LITTELL, R. **Advanced general linear models with an emphasis on mixed models**: course notes. Cary: Statistical Analysis System Institute, 1996. 614 p.

LIMA, M.L.A.; GARCIA, A.A.F.; OLIVEIRA, K.M.; MATSUOKA, S.; ARIZONO, H.; DE SOUZA JR, C.L.; DE SOUZA, A.P. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugarcane (*Saccharum* spp.). **Theoretical and Applied Genetics**, New York, v. 104, p. 30-38. 2002.

LINDSEY, J.K. **Models for Repeated Measurements**. 2th. ed. Oxford University Press, 1999. 536 p.

LITTELL, R.C.; STROUP, W.W.; FREUND, R.J. **SAS for Linear Models**. 4th. ed. Cary: Statistical Analysis Sistem Institute, 2002. 496 p.

LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D. **SAS system for mixed models**. Cary: Statistical Analysis Sistem Institute, 2006. 814 p.

LORENZANA, R.E.; BERNARDO, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical and Applied Genetics**, New York, v. 120, p. 151-161, 2009.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980 p.

MACKILL D.J; NGUYEN H.T; ZHANG, J. Use of molecular markers in plant improvement programs for rainfed lowland rice. **Field Crops Research**, Amsterdam, v. 64, p. 177-185, 1999.

MALÉCOT, G. **The Mathematics of Heredity**. San Francisco: Freeman, 1969. 88 p.

MARGARIDO, G.R.A. **Mapeamento de QTLs em múltiplos caracteres e ambientes em cruzamento comercial de cana-de-açúcar usando modelos mistos**. 2011. 107 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2011.

MARIOTTI, J.A. Experiencias de seleccion clonal em cana de açúcar en la provincia de Jujuy: Repetibilidad y heredabilidad de caracteres de interesse agronómico. **Revista Agronómica Norte Argentina**, Argentina, v.10, p. 61-73, 1973.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. Ed. **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999. p. 205-252.

McCULLOCH, C.E.; SEARLE, S.R. **Generalized, linear and mixed models**. New York: John Wiley & Sons, 2001. 325 p.

McLEAN, R.A.; SANDERS, W.L. Approximating degrees of freedom for standard errors in mixed linear models. In: **Proceedings of the statistical computing section**. New Orleans: American Statistical Association, 1988. p. 50-59.

MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Bethesda, v. 157, p. 1819-1829, 2001.

MEYER, K. Factor-analytic models for genotype $\times$ environment type problems and structured covariance matrices. **Genetics Selection Evolution**, Les Ulis, v. 41, p. 1-11, 2009.

MILLIGAN, S.B.; GRAVOIS, K.A.; BISCHOFF, K.P.; MARTIN, F.A. Crop effects on broad-sense heritabilities and genetic variances of sugarcane yield components. **Crop Science**, Madison, v. 30, p. 344-349, 1990.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. New York: McGraw-Hill, 1974. 564 p.

MOORE, P.H. Integration of sucrose accumulation processes across hierarchical scales: towards developing and understanding of the gene-to-crop continuum. **Field Crops Research**, Amsterdam, v. 92, p. 119-135, 2005.

OAKEY, H.; VERBYLA, A.; PITCHFORD, W.; CULLIS, B.; KUCHEL, H. Joint modelling of additive and non-additive genetic line effects in single field trials. **Theoretical and Applied Genetics**, New York, v. 113, p. 809-819, 2006.

OLIVEIRA, R.A.; DAROS, E.; BESPALHOK FILHO, J.C.; ZAMBON, J.L.C.; IDO, O.T.; WEBER, H.; RESENDE, M.D.V.; ZENI NETO, H. Seleção de famílias de cana-de-açúcar via modelos mistos. **Scientia Agrária**, Curitiba, v. 9, p. 269-274, 2008.

OMAN, S.D. Multiplicative Effects in Mixed Model Analysis of Variance. **Biometrika**, Cambridge, v. 78, n. 4, p. 729-739, 1991.

PASTINA, M.M. **Mapeamento de QTLs e estudo da interação entre QTLs, ambientes e cortes em cana-de-açúcar usando a abordagem de modelos mistos**. 2010. 89 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2010.

PASTINA, M.M.; MALOSETTI, M.; GAZAFFI, R.; MOLLINARI, M.; MARGARIDO, G.R.; OLIVEIRA, K.M.; PINTO, L.R.; SOUZA, A.P.; van EEUWIJK, F.A.; GARCIA, A.A.F. A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. **Theoretical and Applied Genetics**, New York, v. 124, p. 835-849, 2012.

PATTERSON, H.D.; THOMPSON, R. Recovery of interblock information when block sizes are unequal. **Biometrika**, Cambridge, v. 31, p. 100-109, 1971.

PATTERSON, H.D.; SILVEY, V.; TALBOT, M.; WEATHERUP, S.T.C. Variability of yields of cereal varieties in UK trials. **The Journal of Agricultural Science**, Cambridge, v. 89, p. 239-245, 1977.

PAULER, D.K. The Schwarz Criterion and Related Methods for Normal Linear Models. **Biometrika**, Cambridge, v. 85, p. 13-27, 1998.

PAYNE, R.W.; MURRAY, D.A.; HARDING, S.A.; BAIRD, D.B.; SOUTAR, D.M. **An introduction to GenStat for Windows**. 14th ed. Hemel Hempstead: VSN International, 2011. 450 p.

PIEPHO, H.P. Analyzing genotype-environment data by mixed models with multiplicative terms. **Biometrics**, Washington, v. 53, p. 761-766, 1997.

PIEPHO, H.P. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. **Theoretical and Applied Genetics**, New York, v. 97, p. 195-201, 1998.

PIEPHO, H.P.; MÖHRING, J.; MELCHINGER, A.E.; BÜCHSE, A. BLUP for phenotypic selection in plant breeding and variety testing. **Euphytica**, Wageningen, v. 161, p. 209-228, 2008.

PIEPHO, H.P. Ridge regression and extensions for genomewide selection in maize. **Crop Science**, Madison, v. 49, p. 1165-1176, 2009.

POLLAK, E.J.; UFFORD, G.R.; GROSS, S.J. Comparison of alternative models for within-herd genetic evaluation fo beef cattle. **Journal of Animal Science**, Albany, v.45, p. 1010-1014, 1977.

PURBA, A.R.; FLORI, A.; BAUDOUIN, L.; HAMON, S. Prediction of oil palm (*Elaeis guineensis*, Jacq.) agronomic performances using the best linear unbiased predictor (BLUP). **Theoretical and Applied Genetics**, New York, v. 102, p. 787-792, 2001.

RABELO, S.C.; CARRERE, H.; MARIEL FILHO, R.; COSTA, A.C. Production of bioethanol, methane and heat from sugarcane bagasse in a biorefinery concept. **Bioresource Technology**, Essex, v. 102, p. 7887-7895, 2011.

RESENDE, M.D.V. **Análise estatística de modelos mistos via REML/BLUP na experimentação em melhoramento de plantas perenes**. Colombo: Embrapa Florestas, 2000. 102 p. (Documentos, 47)

RESENDE, M. D. V.; THOMPSON, R. Factor analytic multiplicative mixed models in the analysis of multiple experiments. **Brazilian Journal of Mathematics and Statistics**, São Paulo, v. 22, p. 31-52, 2004.

RESENDE, M.D.V.; BARBOSA, M.H.P. Selection via simulated individual BLUP based on family genotypic effects in sugarcane. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, p. 421-429, 2006.

RESENDE, M.D.V. **Genômica Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais**. Colombo: EMBRAPA Florestas, 2008, 330p.

SCHULZ-STREECK, T.; PIEPHO, H.P. Genome-wide selection by mixed model ridge regression and extensions based on geostatistical models. **BMC proceedings**, London, v. 4, p. 1-5, 2010.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, Philadelphia, v. 6, p. 461-464, 1978.

SEARLE, S.R. **Linear models**. New York: John Wiley & Sons, 1971. 532 p.

SEARLE, S.R.; CASELLA, G.; McCULLOCH, C.E. **Variance Components**. New York: John Wiley & Sons, 1992. 501 p.

SEGURA, V.; VILHJÁLMSSON, B.J.; PLATT, A.; KORTE, A.; SEREN, Ü.; LONG, Q.; NORDBORG, M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. **Nature genetics**, New York, v. 44, p. 825-830, 2012.

SLAFER, G.A; ARAUS, J.L; ROYO, C; Del MORAL, L.F.G. Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. **Annals of Applied Biology**, Cambridge, v. 146, p. 61-70, 2005.

SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. **Biometrics**, Washington, v. 57, p. 1138–1147, 2001.

SMITH, A.; CULLIS, B.R.; THOMPSON, R. Exploring variety environment data using random effects models with adjustment for spatial field trends. In: KANG, M.S. (Ed.) **Quantitative genetics, genomics and plant breeding**. CABI Publishing, 2002. 400 p.

SMITH, A.B.; CULLIS, B.R.; THOMPSON, R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. **The Journal of Agricultural Science**, Cambridge, v. 143, p. 449-462, 2005.

SMITH, A.B.; STRINGER, J.K.; WEI, X.; CULLIS, B.R. Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. **Euphytica**, Wageningen, v.157, p. 253-266, 2007.

SO, Y.; EDWARDS, J.W. Predictive ability assessment of linear mixed models in multienvironment trials in corn (*Zea mays* L). **Crop Science**, Madison, v. 51, p. 542-552, 2011.

SOUZA, E.; SORRELLS, M.E. Pedigree analysis of North American oat cultivars released from 1951 to 1985. **Crop Science**, Madison, v. 29, p. 595-601, 1989.

SOUZA JÚNIOR, C.L. **Melhoramento de espécies de reprodução vegetativa**. Publicação didática, Departamento de Genética, Piracicaba: Esalq/USP, 1995. 41 p.

STRINGER, J.K.; CULLIS, B.R. Application of spatial analysis techniques to adjust for fertility trends and identify interplot competition in early stage sugarcane selection trials. **Crop and Pasture Science**, Victoria, v. 53, p. 911-918, 2002.

THOMPSON, R.; CULLIS, B.R.; SMITH, A.B.; GILMOUR, A.R. A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. **Australian and New Zealand Journal of Statistics**, Oxford, v. 45, p. 445-459, 2003.

TRETHOWAN, R.M.; REYNOLDS, M.; SAYRE, K.; ORTIZ-MONASTERIO, I. Adapting wheat cultivars to resource conserving farming practices and human nutritional needs. **Annals of Applied Biology**, Cambridge, v. 146, p. 405-413, 2005.

Van der WERF, J. **Mixed Models for Genetic Analysis**. Armidale: University of New England, 2011. 27 p. Disponível em: <<http://www.vsnr.co.uk>>. Acesso em: 15 dez. 2011.

VERBEKE, G.; MOLENBERGHS, G. **Linear mixed models for longitudinal data**. New York: Springer, 2009. 592 p.

VERBYLA, A.; ECKERMAN, P.J.; THOMPSON, R.; CULLIS, B. The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. **Australian Journal of Agricultural Research**, Melbourne, v. 54, p. 1395-1408, 2003.

WHITE, T.L.; HODGE, G.R. **Predicting Breeding Values with Applications in Forest Tree Improvement**. Dordrecht: Kluwer Academic Publishers, 1989. 363 p.

WEI, X.M.; BORRALHO, M.G. Genetic gains and levels of relatedness from best linear unbiased prediction selection of *Eucalyptus urophylla* for pulp production in southeastern China. **Canadian journal of forest research**, Ottawa, v. 30, p. 1601-1607, 2000.

WOLFINGER, R.D. Covariance structure selection in general mixed linear models. **Communications in Statistics A, Theory and Methods**, Zug, v. 22, p. 1079-1106, 1993.

YPMA, T.J. Historical development of the Newton-Raphson method. **SIAM review**, Philadelphia v. 37, p. 531-551, 1995.

YU, J.; PRESSOIR, G.; BRIGGS, W.; VROH, B. I.; YAMASAKI, M.; DOEBLEY, J.F.; McMULLEN, M.D.; GAUT, B.S.; NIELSEN, D.M.; HOLLAND, J.B.; KRESOVICH, S.; BUCKLER, E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, p. 203-208, 2006.

ZHE-ZHANG.; LIU, J.; DING, X.; BIJMA, P.; DE KONING, D.J.; ZHANG, Q. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. **PLoS One**, Cambridge, v. 5, p. 1-8, 2010.

ZHIWU-ZHANG.; ERSOZ, E.; LAI, C.Q.; TODHUNTER, R.J.; TIWARI, H.K.; GORE, M.A.; BRADBURY, P.J.; YU, J.; ARNETT, D.K.; ORDOVAS, J.M.; BUCKLER, E.S. Mixed linear model approach adapted for genome-wide association studies. **Nature genetics**, New York, v. 42, p. 355-360, 2010.

APÊNDICE

APÊNDICE 1 - Matrizes de VCOV genética para local e colheita

Na diagonal tem-se a variância de cada local informando sobre a presença de heterogeneidade. Fora da diagonal tem-se a covariância entre os pares de locais que informa sobre a presença de interação $G \times L$

Figura 11 – Matriz de variâncias e covariâncias, $\mathbf{G}_{44 \times 44}^L$, do tipo Fator Analítico de Primeira Ordem - FA1, ajustada para efeito genético de genótipos nos 44 locais (g_{ij}), implementada no Modelo 11. Na diagonal tem-se a variância de cada local informando sobre a presença de heterogeneidade. Fora da diagonal tem-se a covariância entre os pares de locais que informa sobre a presença de interação $G \times L$

Figura 12 – Matriz de variâncias e covariâncias, $\mathbf{G}_{5 \times 5}^C$, Autoregressiva de Primeira Ordem - AR1, ajustada para efeito genético de genótipos nas 5 colheitas (g_{ik}), implementada no Modelo 11. Na diagonal tem-se a variância de cada colheita informando sobre a presença de heterogeneidade. Fora da diagonal tem-se as covariâncias entre os pares de colheitas informando a presença de interação $G \times C$ e as correlações genéticas entre as colheitas, que decrescema medida que aumenta a distância temporal entre colheitas